

Result No.	Query %			ID	Description
	Score	Match	Length		
1	1282	74.4	331	1 W06800	Toluene ortho-mono
2	832	48.3	331	1 W98971	Alcaligenes sp. pr
3	179.5	10.4	342	1 R66213	Nocardia corallina
4	178.5	10.4	343	1 R14659	Nocardia corallina
5	86.5	5.0	370	1 R03056	Beta-1,3-glucanase
6	86	5.0	551	1 W18790	Corrected Bacillus
7	85.5	5.0	359	1 R07315	Beta-1,3-glucanase
8	84.5	4.9	337	1 W27608	Tomato PR-gluca
9	84.5	4.9	359	1 R03054	Beta-1,3-glucanase
10	84.5	4.9	359	1 R03055	Beta-1,3-glucanase
11	84.5	4.9	359	1 R14028	Beta-1,3-glucanase
12	84.5	4.9	359	1 R22005	Protein encoded by
13	84.5	4.9	359	1 W83440	Beta-1,3-glucanase
14	84.5	4.9	359	1 W73543	Beta-1,3-glucanase
15	84	4.9	325	1 W82241	Tobacco beta-1,3-g
16	84	4.9	329	1 W87563	A protein with ste
17	84	4.9	330	1 W87564	A protein with ste
18	84	4.9	15281	1 R44929	A protein with ste
19	83.5	4.8	887	1 W77409	T. niveum Cyclospo
20	82	4.8	437	1 W05299	Mycobacterium Kans
21	81.5	4.7	504	1 W98973	Esterase secretory
22	81	4.7	241	1 W36602	Alcaligenes sp. pr
23	81	4.7	506	1 X01313	Streptococcus pneu
24	81	4.7	1194	1 W39272	Glutamine-oxogluta
25	80	4.6	396	1 R76550	Porcine retrovirus
26	79.5	4.6	1721	1 W52847	The most stable alka
27	79.5	4.6	1810	1 R94563	A. mediterranei ri
28	79	4.6	1145	1 W32097	Chicken cytotactin
29	78	4.5	344	1 R10864	Miniature swine re
30	78	4.5	397	1 W93147	Lipase modulating
31	78	4.5	457	1 R86803	Bacillus sp. xylan
32	78	4.5	1611	1 W22504	S. clavuligerus ly
33	76.5	4.4	359	1 R13277	Tylosactone synthase
					Tobacco intracellular

101 LEAAKATWTRDAAWQPLRR

QY 181 LEAAKATWTRDAACWPLRRYYVEDTLVADPVSELFIAQNALDGLIPLYVDREWDERIAL 240

: | | | : | | | | : | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |

Db	172	TWSQDIHLAADNPVSHSNWYVALHFYSGTHGQFLDRIRITYANKGAAL	-----FVIEGT	226
QY	259	-----ESNRWIDAVYKTMAAESDDNFALLARWTRDWS	-----ARAEAAALPAA	302
Db	227	SDASNGGYPYFQSQEWIDFL	-----NARKISWV-NWSLADKVTSAALMPGAS	274
QY	303	-----RALQDAGRAALDEVRE	318	
Db	275	PTGGWTDQAQLSESGKWRDQIRQ	297	

RESULT 7

R07315 ID R07315 standard; protein; 359 AA.

AC R07315; AC

DT 31-JAN-1991 (first entry)

DE Beta-1,3-glucanase plant pathogenesis-related protein.

DE Transgenic plants; disease resistance; chimeric DNA;

KW plant pathogenesis-related protein.

KW synthetic.

OS EP-392225-A.

PN EP-392225-A.

PD 17-OCT-1990.

PF 21-MAR-1990; 105336.

PR 24-MAR-1989; US-329018.

PR 20-JUN-1989; US-368672.

PR 20-OCT-1989; US-425504.

PR (CIBA) CIBA GEIGY AG.

PI Ryaals JA, Alexander DC, Goodman RM, Meins F, Payne GB;

PI Scinson JR, Neuhaus J-M, Moyer MB;

PI WPI; 90-313983/42.

DR N-PSDB; Q06208.

DR Disease-resistant transgenic plants - obt'd. using encoding an

PT inducible pathogenesis-related protein from infected plants.

PS Example 32; page 46; 77pp; English.

PS This is the sequence of the plant pathogenesis-related protein

CC (PRP), beta-1,3-glucanase. It confers systemic acquired resis-

CC tance to plants. The corresp. DNA is used, in a chimeric cons-

CC truct, to produce transgenic plant cells or -tissues with the

CC ability to regenerate into plants which are disease resistant.

CC See also Q06179-86, Q06199-Q06207 and Q06829.

SC Sequence 359 AA;

Query Match	5.0%;	Score 85.5;	DB 1;	Length 359;
Best Local Similarity	22.3%;	Pred. No. 0.48;		
Matches	54;	Conservative 34;	Mismatches 71;	Indels 83;
Gaps				

QY	58	IFDPSGAIRMANWYALKDPQFYVYASWATTRAQODAMESFEFVESRMIGLMRDVVA	117
	:		
	:		
Db	55	LYDPNHGALQ-----ALKG-----	83
		-----SNTEV-----	MLGLPNSDK 83
QY	118	ARALDVLVPLRHAAGANMN-----NAQICALG-----	YGVTFAPAMFHAM 159
	:		:
	:		:
Db	84	HSASG-----MEHARWWQKVKVDFWPDVKIKYIavgNEISPTGTSYLTSFLTTPAMVNIY	139
QY	160	DNLCVAGYILRLALANEPDVL-----EAKATWTRDAAW--QPLRRYVEDT-----LV--	206
	:		:
	:		:
Db	140	KAIGEALGHNRIKVSVDMTLIGNSYPPSQGSFRNDARWFTDPIVGFRLDTRAPILLVNI	199
QY	207	-----YADPVEFLIAQNAL-DG-LLYPLVYDFVDE-KIALE--GGSAVAML 249	
	:		:
	:		:
Db	200	YPFYSYSGNPGQISLFXSLFTAPNVVQDGSQRNYRLFDAMLDSVYAALERSGGASGVIG	259
QY	250	TA 251	
	:		
Db	260	VS 261	
	:		

RESULT	8
W27608	
ID	W27608 standard; Protein; 337 AA.
AC	W27608;

DT 28-MAY-1998 (first entry)
DE tomato PR-glucanase.
KW PR-glucanase; transgenic plants; male sterility.
OS Lycopersicon species.
PN W09738116-A1.
PD 16-OCT-1997.
PF 10-APR-1997; G000992.
PR 11-APR-1996; GB-007517.
PA (GENE-) GENE SHEARS PTY LTD.
PI PAUL W, Perez P, Scott RJ;
DR WFI; 97-512728/47.
DR N-PSDB; T73504.
DR Use of glucanase DNA to prepare male sterile transgenic tomato plant
PT - and complementary sequences for restoring fertility, useful in
PT hybrid seed production
PS Disclosure; Fig 5; 46pp; English.
CC This is PR-glucanase of the tomato plant, and it has areas of
CC complementarity to the ribozyme coding region. The DNA sequence, (or
CC homologue), encoding this protein, (or homologue), operably linked to
CC a promoter or other regulatory sequence (providing suitable tissue and/or
CC temporal specific expression), can be used to prepare a male sterile
CC transgenic tomato plant. The male sterile tomato plants can be used to
CC produce hybrid seeds. The DNA sequence of PR-glucanase can be used to
CC restore male fertility to sterile plants containing glucanases.
SO Sequence 337 AA;

RESULT	9	
R03054	R03054 standard; protein; 359 AA.	
ID	AC	
AC	R03054;	
DT	25-JUL-1990 (first entry)	
DE	Beta-1,3-glucanase from cDNA clones pGL28, pGL30, pGL31, pGL43 and pGL36.	
KW	Beta-1,3-glucanase; papermaking; polysaccharide hydrolase;	
KW	ds.	
OS	Nicotiana sp.	
FT	Key	Location/Qualifiers
FT	peptide	1. .21
FT		/label=prepro leader peptide.
FT		22. .359
PN	protein	
PN	EP-353191-A.	
PD	31-JAN-1990.	
PD	20-JUL-1989; 810555.	
FR	29-JUL-1988; US-353312.	
PA	(CIBA) Ciba Gelgy AG.	
PI	Shinshi H, Wenzler H, Hofsteenge J, Ryals J, Sperisen C;	
DR	WPI; 90-031717/05.	
DR	N-PSDB; Q03249, Q03250, Q03251.	

PT Recombinant DNA encoding beta-1,3-glucanase polypeptide(s) -
PT useful in plant protection and biomass conversion pref. isolated
PT from tobacco.
PS Disclosure: Fig 4-6; 4lpp; English.
CC Beta-1,3-glucanase is involved in plant defence against pathogens and as
CC it hydrolyses polysaccharides, useful in biomass conversion or paper
CC making. The clones allow for transfer of the gene to other species and
CC control of gene expression.
CC Sequence 359 AA;
SQ

Query Match 4.9%; Score 84.5; DB 1; Length 359;
Best Local Similarity 21.9%; Pred. No. 0.62;
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;

QY 58 IFDPSRAIRMANWYALKDPQFYASWATTRARQQDAMESNFVFSRRMIGLRRDVA 117

QY 118 ARALDVLPLRHAAWGANMN-----NAICALG-----YGTVFATAMFHAM 159

QY 160 DNLGVAQYLTRLALAMAEPDVL-----EAKATWTRDAW--QPLRYVEDT-----LV-- 206

QY -----VADPVEIFIAQNIAL-DG-LLYPLVYDFVDE-RTALE--GGSAVAML 249

QY 250 TA 251
:

RESULT 10
R03055

AC	R030357	
DT	25-JUL-1990	(first entry)
DE	Beta-1,3-glucanase	derived from cDNA clones pGL28, pGL30 and pGL31.

	Location/Qualifiers
KW	ds.
OS	Nicotiana sp.
FH	
Key	

FT	
FT	/label=prepro leader peptide.
PN	22. .359
PN	protein
PN	EP-353191-A.

PF 20-JUL-1989; 810555.
PR 29-JUL-1988; US-353312.
DA (CIRA) Giba Geogr AC

DR WPI; 90-031717/05.
DR N-PSDB; Q03252.
DR December 13, 1990.
DR

PS user in plant protection and biomass conversion pref. isolated from tobacco.
PS Disclosure; Fig 7; 41pp: English.

CC it hydrolyses polysaccharides, useful in biomass conversion or papermaking. The clones allow for transfer of the gene to other species and control of gene expression.

Query Match 4.9% Score 84.5 DB 1 Length 359:

Matches	53;	Conservative	35;	Mismatches	71;	Indels	83;	Gaps
QV	58	TFDPSRSATRMANVVAITKDDPQEVVYASWATTTTRAPQDQAMPSNEEVEYESDQMTQMDDQVA	117					

[illegible]

RESULT	11
R14028	
TD	P14028

AC	RI4028;
DT	05-DEC-1991 (first entry)
DE	Beta-1,3-glucanase.

OS *Nicotiana tabacum*.
PN EP-448511-A.
PD 25-SEP-1991.

PR 12-MAR-1990; US-491801.
PA (CIBA) CIBA GEIGY AG.
PI Ryals JA, Gay PB, Ahl Goy PA, Garcia-Olmedo F;

DR N-PSDB; Q13750.
PT Anti-pathogenic comps. - comprise lytic peptide(s) and
PT hydrolytic enzymes

The amino acid sequence codes for the basic beta-1,3-glucanase from Ntabacum, it is truncated at the 5' end and does not comprise the entire signal peptide which is necessary to ensure that the beta-1,3-glucanase is properly targeted to the central vacuole in transgenic plants. The enzyme acts to degrade the cell membranes of plant pathogens and can be used in conjunction with one or more hydrolytic enzymes to improve its anti-pathogenic action. See also 013748-013750.

Query Match	4.9%	Score	84.5	DB	1	Length	359
Best Local Similarity	21.9%	Pred.	No.	0.62			
Matches	53	Conservative	35	Mismatches	71	Indels	83
						Gaps	14

QY 58 IFDPGRSAIRMANWYALKDPQFYASWATTRAQDDAMESNFEFVESRRMIGLRRDDVA 117

```

QY      118  ARALDVLPLRHAAGWGNM-----NAQICALG-----YGVTFETAPAMFHAM 159
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

QY 160 DNLGVAQYLTRLALAMAEFDVL-----EAAKATWTRDAW--QPLRRYVEDT---LV-- 206

QY 207 -----VADPVELETAQNIAL-DG-LLYPLVYDRFVDE-RIALE--GCSAVAML 249

Qy	250 TA 251
----	------------

Db 260 vs 261

PR	24-MAR-1989;	US-329018.
PR	20-JUN-1989;	US-368672.
PR	20-OCT-1989;	US-425504.
PR	07-SEP-1990;	US-580431.
PR	21-DEC-1990;	US-632441.
PR	01-APR-1991;	US-678378.
PR	27-SEP-1991;	US-768122.
PR	06-MAR-1992;	US-848506.
PR	06-NOV-1992;	US-973197.
PR	06-APR-1993;	US-042847.
PR	12-APR-1993;	US-045937.
PR	16-JUL-1993;	US-093301.
PR	13-JAN-1994;	US-181271.
PA	(NOVS) NOVARTIS FINANCE CORP.	
PI	Moyer MB, Payne GB, Ryals JA, Ward ER;	
PI	WPI; 99-059180/05.	
PR	N-PSDB; V81609.	
PR	DNA encoding pathogenesis-related glucanase proteins - useful for	
PT	producing transgenic plants with enhanced disease or pest resistance	
PT	Disclosure; Column 247-248; 169pp; English.	
CC	The present invention describes a DNA molecule encoding a	
CC	pathogenesis-related (PR) protein having beta-1,3-glucanase activity	
CC	selected from PR-2', PR-2'', PR-N, PR-O', PR-O''. Also described	
CC	are: (i) a chimeric gene comprising the above DNA molecule linked to a	
CC	heterologous promoter; (ii) a vector containing the chimeric gene;	
CC	(iii) a host cell containing the chimeric gene; (iv) a transgenic plant	
CC	containing the chimeric gene; and (v) a seed from the transgenic plant	
CC	The DNA molecule is used to produce transgenic plants with enhanced	
CC	disease or pest resistance. The present sequence represents a tobacco	
CC	beta-1,3-glucanase protein for plasmid pGLN17 from the present	
CC	invention.	
SQ	Sequence	359 AA;

Query Match	4.9%;	Score 84.5;	DB 1;	Length 359;
Best Local Similarity	21.9%;	Pred. No. 0.62;	71;	Indels 83;
Matches 53;	Conservative 35;	Mismatches 14;		
QY	58	IFDPSRAIRMANWYALKDPQFYASWATRARQQRDAMESFEFVESRRMIGLMRDDVA	117	
		: : : : : :		
Db	55	LYDPNHGALQ-----ALKG-----SNTIEV-----MLGLPNSDK 83		
QY	118	ARALDVLVPLRHAAGANMN-----NAQICALG-----YGTVFETAPAMFHAM 159		
		: : : : : :		
Db	84	-----HTASGMEHARWVQKNVKDFWPDVKIKYIAGNEISFVTGTSYLSFSLTPAWNIY 139		
QY	160	DNLGVAQYLYLRALAMAEPDVL-----EAAKATWTRDAW-----QPLRRYVEDT-----LV-- 206		
		: : : : : :		
Db	140	KATGEAGLGNNIKVSTSDVMTLIGNSYPPSQSGSRNDARWFDPVIGVFLDRTRAPLLVNI 199		
QY	207	-----VADPVELPFIQNAL-DG-LLYPLGVYDFVDE-RIALE--GGSAVAML 249		
		: : : : : :		
Db	200	YPFYSGNGPCQISLPSYSLTAPNVVQDGSROYNLFDAMLDSVYAALERSGASGVIV 259		
QY	250	TA 251		
Db	260	VS 261		

RESULT	14
ID	W73543
WD	W73543 standard; Protein; 359 AA.
AC	W73543;
DT	05-MAR-1999 (first entry)
DE	Tobacco beta-1,3-glucanase protein clone pGIN17.
KW	chemically regulatable DNA promoter; expression control; pesticide;
KW	herbicide tolerance; beta-1,3-glucanase gene.
OS	Nicotiana acuminata.
PN	US5851766-A.
PD	22-DEC-1998.
PF	31-MAY-1995; 456262.
PR	31-MAY-1995; US-456262.
PA	(NOVS) NOVARTIS FINANCE CORP.

us-09-430-029-3.rag

Wed Sep 27 17:16:53 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 19:19:06 ; Search time 60.65 Seconds
(without alignments)
83.660 Million cell updates/sec

Title: US-09-430-029-3
Perfect score: 1723
Sequence: 1 MTELEKTDIKPLRHFAHV.....ALDEVREQFHARAALRGIAL 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282	74.4	331	1	US-08-319-387-2
2	186	10.8	327	5	5171684-6
3	178.5	10.4	342	1	US-08-499-215-2
4	87	5.0	560	2	US-08-756-317-8
5	86	5.0	551	2	US-09-033-537A-1
6	84.5	4.9	359	1	US-08-181-271A-98
7	84.5	4.9	359	1	US-08-449-315-98
8	84.5	4.9	359	1	US-08-444-803-98
9	84.5	4.9	359	1	US-08-449-043-98
10	84.5	4.9	359	1	US-08-456-265A-98
11	84.5	4.9	359	1	US-08-455-416-98
12	84.5	4.9	359	1	US-08-455-244-98
13	84.5	4.9	359	1	US-08-454-876-98
14	84.5	4.9	359	2	US-08-457-364-98
15	84.5	4.9	359	2	US-08-456-262-98
16	84.5	4.9	359	2	US-08-456-240-98
17	84.5	4.9	359	2	US-08-455-736-98
18	84.5	4.9	359	2	US-08-971-217-98
19	84	4.9	15281	2	US-08-471-119A-2
20	82	4.8	437	2	US-08-620-605D-4
21	82	4.8	464	2	US-09-005-232A-4
22	79.5	4.6	1810	4	PCT-US95-11684-4
23	78.5	4.6	516	2	US-09-019-201A-3
24	78	4.5	397	2	US-08-282-197C-55
25	78	4.5	457	1	US-08-206-006-2
26	78	4.5	1611	2	US-08-804-227C-5
27	77	4.5	694	2	US-08-895-522-3
28	77	4.5	694	3	US-09-195-391-3

29	76.5	4.4	359	1	US-08-047-413-15	Sequence 15, Appl
30	76.5	4.4	359	3	US-08-229-050-15	Sequence 15, Appl
31	76.5	4.4	3724	2	US-08-804-227C-10	Sequence 10, Appl
32	76.5	4.4	3724	2	US-08-804-198-4	Sequence 4, Appl
33	76	4.4	471	2	US-08-466-583-2	Sequence 2, Appl
34	76	4.4	471	4	PCT-US95-07820-2	Sequence 2, Appl
35	75.5	4.4	4472	2	US-08-804-227C-2	Sequence 2, Appl
36	75	4.4	559	2	US-08-756-317-10	Sequence 2, Appl
37	74.5	4.3	412	2	US-08-851-088-12	Sequence 10, Appl
38	74	4.3	747	2	US-08-895-522-1	Sequence 12, Appl
39	74	4.3	747	3	US-09-195-391-1	Sequence 1, Appl
40	73.5	4.3	262	2	US-08-602-359A-38	Sequence 38, Appl
41	73.5	4.3	269	1	US-07-706-691G-50	Sequence 50, Appl
42	73.5	4.3	269	1	US-08-254-021-50	Sequence 50, Appl
43	73.5	4.3	269	2	US-08-618-446-50	Sequence 50, Appl
44	73.5	4.3	4545	2	US-08-804-227C-14	Sequence 14, Appl
45	73	4.2	268	2	US-08-835-099A-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-319-387-2
; Sequence 2, Application US/08319387
; Patent No. 5543317
; GENERAL INFORMATION:
; APPLICANT: Shields, Malcolm S.
; APPLICANT: Francesconi, Stephen C.
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene, Dichloroethylenes and Aromatic Pollutants
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/319,387
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/167,457
; FILING DATE: 15-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,718
; FILING DATE: 02-MAY-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UWF-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-319-387-2


```

; Sequence 8, Application US/08756317
; Patent No. 584984
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,693
; FILING DATE: 29-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fatterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MORT:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 787-1440
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-756-317-8

```

```

Query Match          5.0%; Score 87; DB 2; Length 560;
Best Local Similarity 21.8%; Pred. No. 0.39;
Matches 75; Conservative 43; Mismatches 120; Indels 106; Gaps 21;

```

```

QY 5 LKTVDKLRLHTFAHVAONIGDKTATRYOEGMGGAQPOENPHYRTPWDPDYEIFD--PS 62
DB 201 LELIQKPL-----GER---QYAKPLLVPPQINKY-----IFDLSP 236
QY 63 RSAIRMANWYALKDPROFYASWATTRAQOQDAMESNEFFVESRMIGLMDRDVRAALD 122
DB 237 KSFVQ---YALKNNLQVFIW-----RNPDA-----QHREWGL---STYVEALD 275
QY 123 VLVLPLRHAAGA-NMNNQAICGALGYTVFTAPAMF-----AMDNLGVAQYLTRL----- 171
DB 276 QAIEVSREITGSRSVNLACACA---GGILVAALHGLQVRRLRKVSSTYLVLSLDQM 332
QY 172 ---ALAMAEPPDVEAAGA-----TWTR---DAAMQPLRRYVVEDTLVVAD 209
DB 333 ESPALFADAEQTLSSKRSRYGHVLDGRDMAKVFAMRPNDLIW---NYWNNYLLGRQ 389
QY 210 PVELEIA---QNLALDGLLPLVYDRFVERI-----ALE-GGSAVAMLTAFMPWHHTS 260
DB 390 PPAFDILYNNNDNTLPAAFHGELDLFKHNPRLTRPGALEVSGTAVDLGKVAIDSFHVAG 449
QY 261 -----NRWIDAVVKTMAAESDDNRALLARWTRDWSARAEALAP 299
DB 450 ITDHTPW-DAVYRSALLIGQRRFILSN-----SGHIQSILNP 487

```

RESULT 5

```

US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Annette
; APPLICANT: Toft, Merette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

```

```

Query Match          5.0%; Score 86; DB 2; Length 551;
Best Local Similarity 18.6%; Pred. No. 0.49;
Matches 49; Conservative 42; Mismatches 78; Indels 94; Gaps 13;

```

```

QY 132 WGNMNNQAICGALGYTVFTAPAM-----FHMDNLGVAQYLTRLALAMAEPPDVEA 184
DB 53 WGINFRAMRYTAEDGYI-TDFSVKNKVKEAQASIDLGLYIIDHWHILSDGNPNTRYKAQ 111
QY 185 KATWTRDAA-----WQPLRRYVVEDTL-----VVADP 210
DB 112 SKAFFQEMATLYGNTPNVIYEIANEPNGVSNADVKSYAEVITAIKRAIDPDGVVIVGSP 171
QY 211 -----VELFIAONLALDGLLYPL-----VYDRFVERI--ALEGGSAVAMLTAFMPWHHT 258
DB 172 TWSQDIHLAADNPVSHSNVYALHFYSGTHGQFLDRITYAMNKGAAL-----FVTEWGT 226
QY 259 -----ESNRWIDAVVKTMAAESDDNRALLARWTRDWS-----ARAEALAPVNA 302
DB 227 SDASNGGGPYFPQSKWEWIDFL-----NARKISWV-NWSLADKVETSAAIMPAS 274
QY 303 -----RALQDAGRAALDEVRE 318

```


APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/Pl/CSG 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-449-315-98
Query Match 4.9%; Score 84.5; DB 1; Length 359;
Best Local Similarity 21.9%; Fred. No. 0.37;
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;
QY 58 IFDPSRAIRMANWYALKDPQFYVSWATTRARQODAMESFEVESRRMIGLMRDVVA 117
DB 55 LYDPNHALQ-----ALKG-----SNIEV-----MLGLPNSUVK 83
QY 118 ARALDVLPLRHAAGANN-----NAQICALG-----YGVFTAPAMFHAM 159
DB 84 ----HIASGMEHARWVQKNVKDFWPDVKIKYIAGVNEISPVGTISLTSPAMVNTY 139
QY 160 DNLGVAQYLTRLALAMAEEDVL-----EAKATWTRDAW---QPLRRYVEDT-----LV-- 206
DB 140 KAIEAGLGNKIKVSTVDMTLIGNSYPPSGSFRNDAEFVDPVIGFLRDTRAPLLVNI 199
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDFRVE-RIALE--GGSAVAML 249
DB 200 YPFYSYSGNPGQISLPYSLFTAPNVVQVDSRQYRNLFDMLDVYVALERSGGASGVIV 259
QY 250 TA 251
DB 260 VS 261
RESULT 8
US-08-444-803-98
Sequence 98, Application US/08444803
Patent No. 5654414
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08444,803

```

Query Match          4.9%; Score 84.5; DB 1; Length 359;
Best Local Similarity 21.9%; Pred. No. 0.37;
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;

58  ITPSRSAIRMANWVALKDDPROFYASWATTHARQODAMESNFEEVSRRMIGLMRDVA 117
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
55  LYDPNHGALQ-----ALKG-----SNIEV-----MLGLPNSDK 83

118  ARALDVLPLRIAGANNM-----NAQICALG-----TGTVTATAPMHAM 159
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
84  ----HIASGMEHARWVKQNKDFDPDKIKYIAVNGEISPTGTSYLTSLTAPMNIY 139

```



```
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-456-265A-98

Query Match      4.9%; Score 84.5; DB 1; Length 359;
Best Local Similarity 21.9%; Pred. No. 0.37;
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;

QY 58 IFDPSRAIRMANVYALKDPQFYFASWATTRARQQDAMESNFEVESRRMTGLMRDDVA 117
Db 55 LYDPNHGALQ-----ALKG-----SNIEV-----MLGLPNSDVK 83

QY 118 ARALDVLPLRHAAGANNM-----NAQICALG-----YGTVFTAPAMFHAM 159
Db 84 ----HIASGMEHARWVKQKVKDEWPDVKIKYIAGNEISPVTGTSYLTSLTPAMVNIY 139
QY 160 DNLGVAQYLTLALAMRPDVL-----EAKATWTRDAW--QPLRRYVEDT-----IV-- 206
Db 140 KAIGEAGLGNKIKYSTVDMTLIGNSPQSGSFNDARFVDPITVGFLRDTAPLLVNI 199
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDFVDE-RIALE--GGSAAVAML 249
Db 200 YPFYSYSGNPGQISLPLTAPNVVQDGSQRNLFEDAMLDSVYAALERSGGSAGVIV 259
QY 250 TA 251
Db 260 VS 261

RESULT 11
US-08-455-416-98
; Sequence 98, Application US/08455416
; Patent No. 5777200
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neubaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESS: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/455,416
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-455-416-98

Query Match      4.9%; Score 84.5; DB 1; Length 359;
Best Local Similarity 21.9%; Pred. No. 0.37;
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;

QY 58 IFDPSRAIRMANVYALKDPQFYFASWATTRARQQDAMESNFEVESRRMTGLMRDDVA 117
Db 55 LYDPNHGALQ-----ALKG-----SNIEV-----MLGLPNSDVK 83

QY 118 ARALDVLPLRHAAGANNM-----NAQICALG-----YGTVFTAPAMFHAM 159
Db 84 ----HIASGMEHARWVKQKVKDEWPDVKIKYIAGNEISPVTGTSYLTSLTPAMVNIY 139
QY 160 DNLGVAQYLTLALAMRPDVL-----EAKATWTRDAW--QPLRRYVEDT-----IV-- 206
Db 140 KAIGEAGLGNKIKYSTVDMTLIGNSPQSGSFNDARFVDPITVGFLRDTAPLLVNI 199
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDFVDE-RIALE--GGSAAVAML 249
Db 200 YPFYSYSGNPGQISLPLTAPNVVQDGSQRNLFEDAMLDSVYAALERSGGSAGVIV 259
QY 250 TA 251
Db 260 VS 261
```



```
Db 84 -----HIASGMEHARWVQKVKDFWPDVKIKYIAGVNEISPVIGTSYLTSLTPAMVNIY 139
QY 160 DNLGVAQYLTRLALAMAEFDVL-----EAAKATWTRDAW--OPLRRYVEDT----IV-- 206
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 140 KAIGEAGLGNKIKVSTVDMTLIGNSYPPSGSFRNDARWFVDPIVGFRLDTRAPLVNI 199
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDRFVDE-RIALE--GGSAVAML 249
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 200 YPFYSYGNPGQISLPSYSLFTAPNVVVQDGSQRQYRNLFDAMLDSVYAALERSGGASGV 259
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 250 TA 251
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 260 VS 261

RESULT 12
US-08-455-244-98
; Sequence 98, Application US/08455244
; Patent No. 5789214
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,244
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/PL/CSC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-455-244-98
```

```
Query Match 4.9%; Score 84.5; DB 1; Length 359;
Best Local Similarity 21.9%; Pred. No. 0.37;
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;
```

```
QY 58 IFDPSRAIRMANWYALKDPQFYFVYASWATTRARQQDAMESNFVFESRRMIGLMRDVYA 117
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 55 LYDPNHEGALQ-----ALKG-----SNIEV-----MLGLPNSDVK 83
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 118 ARALDVLVPLRHAANGANN-----NAQICALG-----YGVFTAPAMFHAM 159
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 84 ----HIASGMEHARWVQKVKDFWPDVKIKYIAGVNEISPVIGTSYLTSLTPAMVNIY 139
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 160 DNLGVAQYLTRLALAMAEFDVL-----EAAKATWTRDAW--OPLRRYVEDT----LV-- 206
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 140 KAIGEAGLGNKIKVSTVDMTLIGNSYPPSGSFRNDARWFVDPIVGFRLDTRAPLVNI 199
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDRFVDE-RIALE--GGSAVAML 249
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 200 YPFYSYGNPGQISLPSYSLFTAPNVVVQDGSQRQYRNLFDAMLDSVYAALERSGGASGV 259
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 250 TA 251
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 260 VS 261
```

```
RESULT 13
US-08-454-876-98
; Sequence 98, Application US/08454876
; Patent No. 5804693
; GENERAL INFORMATION:
```


US-08-457-364-98

QY 58 IFDPSRAIRMANWYALKDPROFFYASWATTRAROODAMESNFEFEVSRRMITGLMPDDYA 117

55 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 104

CO. MAGNETIC

[illegible]

LD 84 ----HIASGMEHARWWVQKNVKDFWPDVKIKYIAVGNEISPVGTGYLTSELTSPAMVNTY 139

QY 160 DNLGVAQYLTRLALAMAEFDVL-----EAKATWTRDAW--QPLRRYVEDT----LV-- 206

Db 140 KAIGEAGLGNNIKVSTSVDMTILIGNSVPPSOGSERNDAPWENTPTVCEP POMPADITANT 100

THE UNIVERSITY OF CHICAGO

[illegible]

220 IFFISGNFQIISLFIAPFNVVVQDGSRQYRNLFEDAMLDVYAALERSGGASVGIV 259

QY 250 TA 251

Db 260 VS 261

RESULTS 15
PS-09-156-263-00

; Sequence 98, Application US/08456262

TELETYPE NO: 3831700
; GENERAL INFORMATION:

APPLICANT: Ryals, John A.

APPLICANT: Beck, James J.

APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.

APPLICANT: Goodman, Robert M.

APPLICANT: Meins, Jr., Frederick

APPLICANT: MONCOYA, ALICE
APPLICANT: Mover, Mary R

APPLICANT: Neuhaus, Jean-Marc

APPLICANT: Sperison, Christoph

APPLICANT: STANSON, JEFFREY R.
APPLICANT: UKNES, SCOTT J.

APPLICANT: Ward, Eric R.

; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

7 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106

; CORRESPONDENCE ADDRESS:

STREET: 7 Skyline Drive

CITY: HAWAII
STATE: New York

COUNTRY: USA

; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```

; CURRENT APPLICATION DATA:
; CURRENT RELEASE #1.07 VERSION #1.23

```

APPLICATION NUMBER: US/08/456,262
FILING DATE: 31-MAY-1995

CLASSIFICATION: 435

APPLICATION NUMBER: 08/181,271

FILING DATE: 13-JAN-94
APPLICATION NUMBER: 08/003 201

```

Db      200 YPYFSYSGNFQISLPYSLFTAPVVVQDGSROYRNLFDAMLDSVYAALERSGGASGVIV 259
Qy      250 TA 251
        :
Db      260 VS 261

Search completed: September 26, 2000, 20:22:37
Job time: 3811 sec

```

Query Match	4.98;	Score	84.5;	DB	2;	Length	359;
Best Local Similarity	21.98;	Pred. No.	0.37;				
Matches	53;	Conservative	35;	Mismatches	71;	Indels	83;
Gaps	14;						
QY	58	IFDPSRAIRMANVYALKDPQFYASWATTARQQDAMESNFEEVERMIGLMRDVA	117				
Db	55	LYDENHGAQLQ-----ALKG-----	83				
QY	118	ARAFLDVLPLRHAAGANN-----NAQICALG-----	159				
Db	84	---HIASGMGEHARWVQKNVKDFPDWKIKYIAGNIEISPTGTSYLTSELTAMVNIY	139				
QY	160	DNLGVAOYLTRLATAEPDVL-----EAAKATWTRDAAW--QPLRRYVEED----	206				
Db	140	KAIGEAGLGNNIKVSTSDVMTLIGNSPYSGSFFNDARFWFVDFIVGELRDRAPLLVNI	199				
QY	207	-----VADPVELFIAQNAL-DG-LLXPLVYDREVDE-RIALE--GGSAVAML	249				

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 19:23:36 ; Search time 82.1 Seconds
(without alignments)
249.499 Million cell updates/sec

Title: US-09-430-029-3
Perfect score: 1723
Sequence: 1 MTELEKTVDIKPLRHTFAHV.....ALDEVREQFHARAARLGIAL 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_64:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	779.5	45.2	331	1 B37831	phenol 2-monooxyge
2	778.5	45.2	331	1 S44304	phenol 2-monooxyge
3	640.5	37.2	333	1 S47288	phenol 2-monooxyge
4	123	7.1	394	2 S5208	methane monooxygen
5	105.5	6.1	1092	2 T18305	replication factor
6	105.5	6.1	1092	2 T18306	replication factor
7	105	6.1	505	2 S39962	endoglucanase - Er
8	103	6.0	387	2 J10101	methane monooxygen
9	99	5.7	580	2 G70763	probable maltoolig
10	97.5	5.7	8563	2 T30226	polyketide synthas
11	96	5.6	1238	2 T03465	probable exonuclea
12	91.5	5.3	472	1 G70932	probable monooxygen
13	90.5	5.3	697	2 C75525	molybdopterin oxid
14	90.5	5.3	875	2 G75442	ATP-dependent Clp
15	90	5.2	539	2 A58538	4-hydroxybenzoate-
16	89.5	5.2	415	2 B70898	probable exonuclea
17	89.5	5.2	522	2 A75412	conserved hypothet
18	88.5	5.1	467	2 S58233	PopQ protein - Rhi
19	88.5	5.1	474	2 G72658	probable Vir B11 A
20	88.5	5.1	504	2 S54744	cellulase (EC 3.2.
21	88.5	5.1	509	2 T45408	pet112-like protei
22	88	5.1	1309	2 T00078	probable RNA-direc
23	87.5	5.1	463	2 S01984	hypothetical prote
24	87.5	5.1	1147	2 T35781	hypothetical prote
25	87.5	5.1	1396	2 S36851	L-shaped tail fibe
26	87	5.0	560	2 S29307	poly(3-hydroxyalka
27	87	5.0	573	2 T10037	hypothetical prote
28	87	5.0	1240	2 T04193	hypothetical prote
29	86.5	5.0	454	2 T46818	probable monooxyge

30	86.5	5.0	555	2 F65254	ABC transporter in
31	86.5	5.0	1273	2 E72611	probable ATP-depen
32	86.5	5.0	3158	3 T17483	peptide synthetase
33	85.5	5.0	474	2 JX0301	ubiquinol--cytochr
34	85.5	5.0	509	2 D70856	phosphoribosylam
35	85.5	5.0	526	2 A81135	O-antigen biosynth
36	85.5	5.0	1275	2 T18556	glucan endo-1,3-be
37	85	4.9	369	2 S12406	rec protein - Stre
38	85	4.9	456	2 A31844	hydroxyneurospore
39	85	4.9	485	2 S32171	probable fadE6 pro
40	85	4.9	731	2 B70628	glucan endo-1,3-be
41	84.5	4.9	359	2 A30758	1,3-beta-glucanase
42	84.5	4.9	365	2 S13594	glucan endo-1,3-be
43	84.5	4.9	370	2 A39115	glucan endo-1,3-be
44	84.5	4.9	370	2 B39115	endo-1,4-beta-xyla
45	84.5	4.9	506	1 S59634	

ALIGNMENTS

RESULT 1

B37831

phenol 2-monooxygenase (EC 1.14.13.7) chain P1 - Pseudomonas sp. (strain CF600)
C:Species: Pseudomonas sp.

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: B37831

R:Nordlund, I.; Powlowski, J.; Shingler, V.

J. Bacteriol. 172, 6828-6833, 1990

A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phen

A:Reference number: A37831; MUID:91072230

A:Accession: B37831

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-331 <NOR>

A:Cross-references: GB:M60276; GB:M37764; MID:gl51449; PIDN:AAA25940.1; PID:gl51451

C:Superfamily: phenol 2-monooxygenase component L

C:Keywords: oxidoreductase

Query Match 45.2%; Score 779.5; DB 1; Length 331;
Best Local Similarity 46.8%; Pred. No. 4.7e-59;
Matches 155; Conservative 54; Mismatches 115; Indels 7; Gaps 4;

QY	1	MTELEKTVDIKPLRHTFAHVAQNI	GDKTATRYQEGMGAQPOENH	HYRETWDPDEYIFD	60
Db	1	MSVEIKTNIVDPTRQTYGNLQRRF	-GDKFASRYQEAASYDIEAVTN	PHYRLWDPQHELHD	59
QY	61	PSRSATRMANWYALKDP	PROFYFASWATTRARQODAMES	FEFVESRRMIGLMRDDVAARA	120
Db	60	PIRTALRMTDHWKVTDP	PROFYFAYQVYQTRARQGEATE	HAYGFCEKRELLSRLPAETQAKL	119
QY	121	LDVLVPLRHAANGANNNAQIC	ALGYGVTFATPAMPHANDNLG	VAQYLTLALAM--AEP	178
Db	120	LRLVPLRHAELGANNNSI	AGDSIAATVQMHYQAMDRLG	MGQYLSRIGLLDGGTG	179
QY	179	DVLEAKAKATWTRDAANQ	PLRRYVEDTLVADPVELL	IAQNLALDGLLYPLVYDRFDERI	238
Db	180	EALDQAKAYWLDPITWQ	LRRYVEDSFVDMFELG	IAQNLVLDGLQLPLMYQRF--DQWL	238
QY	239	ALEGGSAVALMTAFMP	EWHTESNRWIDAVVKTMAAES	DDNRALLARWTRGWSARAALA	298
Db	239	TENGSGSDVAMLI	TEFMDKDWGSETRWYDAMF	KTIVLAENDANREQVQAWLEWTEPRATEALL	298
QY	299	PVAARALQDAGRAAL	DEVREQFHARAARLGI	329	
Db	299	PLAEEA---TGIAAL	DEVRSFAFATRLQKTGL	326	

RESULT 2

S44304

phenol 2-monooxygenase (EC 1.14.13.7) component L - Pseudomonas putida

N:Alternate names: phenolhydroxylase chain B

Db 47 KPLSEYQSLSCYAPNPWTAGLDWGDWTKQFGGRPSWGN-----STELRTTDWY 99
QY 73 ALKDP-RQFY--YASWATTRARQADAMESNEFEVESRRMIG-LMRDDVAARALDVLPLR 128
Db 100 RHRDPARWHHPYKDKSEARYQRELAAYSSSGSIRTDIPYRDEILNKYFALLYSE 159
QY 129 HAAGANNNAQICALGYGVFTAPAMFHAMDNGLGVAQ--YLRALALAMAEPOV---LEA 183
Db 160 YGLFNAHSSVGRDC--LSDFIRQTAVEAALDKVDNAQMIQMERLFTAKLVPGFEDASDV 216
QY 184 AKATWTRDAAQWPLRRYVEDTLV-VADVELFIAQNALDGLLPLVYDVRVDERIALEG 242
Db 217 PKIITTDPIYSGARATVQEIWOGVODWNEILWAGHAVMIATFGQFARREFFORLATVYG 276
QY 243 GSAVAMUTAFMPEWHITESNRWIDAV-VKTMAAESD---DNEALLARWTRDWSARAALA 298
Db 277 DTLPPFTAQSTVPTQTRGAIIDLTVYCLANDSEFGAHHNRTFLNMTHEHYLASSVAALK 336
QY 299 P-VAARALQDAGRAALDEVREQFHARAARLGI 329
Db 337 DFVGLYAKVEKSA--DRSRRRLRGAAASSAI 366

RESULT 5
Tl8306
replication factor C like protein - Emericella nidulans
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: Tl8305
R:Kafer, E.; May, G.
Gene 191, 155-159, 1997
A:Title: The uvvF gene region in Aspergillus nidulans codes for a protein with homology
A:Reference number: Z18873
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1092 <KAF>
A:Cross-references: EMBL:U86619; NID:g1840141; PID:g1840142; PIDN:AAB63523.1
C:Genetics:
A:Gene: uvvF

Query Match 6.1%; Score 105.5; DB 2; Length 1092;
Best Local Similarity 20.2%; Pred. No. 0.62;
Matches 69; Conservative 49; Mismatches 144; Indels 79; Gaps 13;

QY 26 GDKT----ATRYQEGMGAQPOENFHYRPTWDPDYEIFDPSSAIRMANWALKDPROFY 81
Db 705 GDQTGDYSLDYDQGMKAWEKHIILKPWDIVGKILNP--QMFSQSSTATLNDKSELY 762
QY 82 -----YASWATTRARQADAMESNEFEVESRRMIGLMDIDVAAR-----ALD 122
Db 763 FNDHEFSYLMQENLYRTKPTRANNYEGKERKLLKLE-----LADNAASSISDGLDLD 815
QY 123 VLPLRHAAGANNMNA-----QICALGYGVFTAPAMFHAMDNGLGVAQYLRALALAMAE 177
Db 816 RMIHCSQQQWSLMPHTAIFSVPRPASFVFGNMMPERPAFTSWLGQNSKGLIIVGVIARE 875
QY 178 PDVLEAKATWTRDAAQWPLRRYVEDTLVADPVELFIAQNALDGLD-----LILYPLIVY 230
Db 876 ADVVHAGK-----LSRYVK-----EIQGRMLRSLRSGDRHVRQOYIPLIW 915
QY 231 DRFYDERTALEGGSAVAMLTAFMPEWHITESNRWIDAVVKTMAESDDNRALL-----ARW 285
Db 916 DRLV-KRLMWDGKSDVEDVDFMDSYFTRDDW-DALVELGLGPMDDSKVKLETQTKAAF 973
QY 286 TRDWSAR-----AEALAPVAARALQDAGRAALDEVRE 318
Db 974 TRVYNQRSHPLPYMKASNLAPKAPKEKPDIEDAIDESDE 1014

RESULT 6
Tl8306
replication factor C protein - Emericella nidulans
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: Tl8306
R:Kafer, E.; May, G.
Gene 191, 155-159, 1997
A:Title: The uvvF gene region in Aspergillus nidulans codes for a protein with homology
A:Reference number: Z18873
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1092 <KAF>
A:Cross-references: EMBL:U86620; NID:g2258283; PIDN:AAB63574.1
C:Genetics:
A:Introns: 29/3; 38/1; 68/2
A>Note: uvvF

Query Match 6.1%; Score 105; DB 2; Length 505;
Best Local Similarity 23.2%; Pred. No. 0.25;
Matches 68; Conservative 38; Mismatches 73; Indels 114; Gaps 17;

QY 112 MRDDVAARALDVLPLRHAAGANNMNAQI--CALGYGVFTAPAMFH-----AMNLL 162
Db 81 LRDD-----WGINVFRVMTAADGY---ISNPSLANKVEAAVAAQSL 121

A;Residues: 1-8563 <APA>
A:Cross-references: EMBL:X86780; NID:q987088; PID:q987100; PTDN:CAA60460.1

C;Genetics:

A;Gene: DR0397

A;Map position: 1

C;Superfamily: formate dehydrogenase

Query Match 5.3%; Score 90.5; DB 2; Length 697;
Best Local Similarity 20.7%; Pred. No. 6.6;
Matches 64; Conservative 32; Mismatches 106; Indels 107; Gaps 15;

QY 60 DPSRAITMANWYALKDPRQFYYSWATTRARQDAMESNFE--FVESRR----- 107
DB 53 DGSEKAVKLIG-----DPAHYTRTGFCAKTVHYPARONHPERPLPLRYNPKTPDPEPQ 107
QY 108 -----MIGLMRDDVAARALDVLPLR-----HAAW--GAMMNA 139
DB 108 WERTWDEALDDIAGRLQLLAERGPQSILPHYAGNMGLEGSHVHALWRALGAELLEE 167
QY 140 QICA-----LGYGVFTTAPAMF--HA-----MDNLGVAQYLTALAMAPDVYLE 182
DB 168 TICASAGTAANEVGYGTRLAVDPLDVPARLIVLWGLNSLTHSLT-----PQI-T 218
QY 183 AAKATWTRDAWQPLRRYVETLVVADPVELFIAQNLALDGLLYPLVYDRFVDERIALEG 242
DB 219 AARKNGAR-----VVCDFYRNRTA-----AFADHLKIRP 249
QY 243 GSAVAMLTAFMPWHITESNRWIDAVVKTMAAES-DDNEALLARWTRDWSARAEALAPVA 301
DB 250 GTDRAALGVWRELFV--NGWTDAEYLAQATEGVDDLR--AAEWSPORTAEVTGLT 303
QY 302 ARALQDAGR 310
DB 304 ADEVROFAR 312

RESULT 14

G75442

ATP-dependent Clp proteinase, ATP-binding subunit ClpB - Deinococcus radiodurans (strain

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C;Accession: G75442

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896

A;Accession: G75442

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-875 <WHI>

A;Cross-references: GB:AE001956; GB:AE000513; NID:G6458776; PIDN:AAF10620.1; PID:G645877

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR1046

A;Map position: 1

C;Superfamily: endopeptidase Clp ATP-binding chain

Query Match 5.3%; Score 90.5; DB 2; Length 875;
Best Local Similarity 23.6%; Pred. No. 8.9;
Matches 78; Conservative 37; Mismatches 109; Indels 107; Gaps 17;

QY 76 DPRQFYASWATTRARQDAMES-NFEVESRKMIGLMRDDVAARALDVLPLRHAAWGA 134
DB 25 NPERFTASAVAINAAQOLAQENHONLTHFVLRTLDNDTASRALTLA-----GG 76
QY 135 NMN-----NAQICAL-----GYGVFTTAPAM--FHAMNLG-----VAQYLTRIAL 173
DB 77 DLNTRAAALDAEIAKLPRKVGSDGQLYLDPALNRAFAQADTLAAQLGDSFVAADTLLEAL 136
QY 174 -----AMAEPPVLEAKATW-----TFDAWQP-- 196

DB 137 RGEYRGRLPDEVSLNRAVTEQRKGKTVINKTSEQFDALNKYGTDLTQARDEGKFDPIV 196
QY 197 -----LRRYVEDTLVADP--VELFIAQNLAL-----DGLLYPLVYDRF 233
DB 197 GRDEIRRMVQILLRRSKNNPNVLIGEIPGVGKTAIAEGLAMRIVKGDVPEGLR----- 248
QY 234 VDERT-ALEGGSAVAMLTAFMPWHITESNRWIDAVVKTMAAE-----SDNRALLARWTRD 288
DB 249 -DKKIVSLEMSSLLAG-AKFRGEFEERLKGVIDEVVKS-AGEIILFVDEIHTIVGAGKTE 305
QY 289 WSAAREALAPVAARA-LQDAGRAALDEVRE 318
DB 306 GSPDAGNMLKPALARGELHLIGATTILDEVRE 336

RESULT 15

A58538

4-hydroxybenzoate--CoA ligase (EC 6.2.1.27) - Rhodopseudomonas palustris

C;Species: Rhodopseudomonas palustris

C;Date: 31-Mar-1997 #sequence_revision 11-Apr-1997 #text_change 18-Feb-2000

C;Accession: A58538; A36960

R;Egland, P.; Harwood, C.S.

submitted to the Protein Sequence Database, March 1997

A;Description: Rhodopseudomonas palustris 4-hydroxybenzoate coenzyme A ligase precurs

A;Reference number: A58538

A;Contents: correction

A;Accession: A58538

A;Molecule type: DNA

A;Residues: 1-539 <EGI>

A;Cross-references: GB:U02033; NID:G665919; PIDN:AAA62604.1; PID:G665920

R;Gibson, J.; Dispensa, M.; Fogg, G.C.; Evans, D.T.; Harwood, C.S.

J. Bacteriol. 176, 634-641, 1994

A;Title: 4-hydroxybenzoate-coenzyme A ligase from Rhodopseudomonas palustris: purific

A;Reference number: A36960; MUID:94131941

A;Accession: A36960

A;Status: significant sequence differences

A;Molecule type: DNA

A;Cross-references: GB:U02033; NID:G665919

C;Genetics:

A;Gene: hbaA

C;Superfamily: human SA protein; acetate--CoA ligase homology

C;Keywords: acid-thiol ligase; aromatic hydrocarbon catabolism

F:54-509/Domain: acetate--CoA ligase homology <ACL>

Query Match 5.2%; Score 90; DB 2; Length 539;

Best Local Similarity 20.8%; Pred. No. 5.2;

Matches 65; Conservative 32; Mismatches 93; Indels 122; Gaps 16;

QY 107 RMIGLMRDDVAARALDVLPLRHAAGANMNAQICALGYGT----- 148
DB 198 RRIGYREDDVVFSAAKLFF-----AYG--LGNAMFCPMGIGATSVLYPERPTADSVFDTL 250
QY 149 -----VFTAPAMPHAM-----DNLGVAQYLTALAMAEF----- 178
DB 251 RLHQPTLLFAVPTLYAAMLADPFRSRTETLPDL-----RLCVSAGEPLFAQVGLNWRN 303
QY 179 ---DVLAAKATWTRDAWQPLRRYVEDTLVWADPVELFIAQNLALDGLLYPLVYDR-- 232
DB 304 RFGHDIVNGVGVSTENGHLFLNPLPHAVE-----YGTSGVPVDGYRLRLYDGRGQ 352
QY 233 -FVDERIA---LEGGSAVA-----MLTAFMPWHITESNRW---IDAVVKTMAAESDD 277
DB 353 DVADEIGELLVSGSSAAGYWNQRDKTRTFVGVGTRTGDKYHRRADGVY-TYCGRTDD 411
QY 278 NRALLARWTRDWSAR-----AEALAP-----VAARALQDAGRAAL 313
DB 412 IFKVSIGIWNVSFPETEQALMSHAKVLEAAVIAEADTDGLIKPKAFIVLASRGDIDFG-ALF 470
QY 314 DEVREOFHARAA 325
DB 471 DELKE--HVKSA 480

Search completed: September 26, 2000, 20:24:12
Job time: 3636 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:22:56 ; Search time 58.11 Seconds
(without alignments)
176.517 Million cell updates/sec

Title: US-09-430-029-3

Perfect score: 1723

Sequence: 1 MTEIKTVDIKPLRTFFAHV.....ALDEVREQFHAPAAALGIAL 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	779.5	45.2	331	1	DMPL_PSESP
2	184	10.7	326	1	TMOE_PSEME
3	123	7.1	394	1	MEMB_METTR
4	106	6.2	444	1	GUNN_ERWCA
5	105	6.1	505	1	GUNV_ERWCA
6	103	6.0	386	1	MEMB_METCA
7	99	5.7	580	1	YF62_MYCTU
8	91.5	5.3	472	1	Y568_MYCTU
9	88.5	5.1	461	1	KLB2_ECOLI
10	88.5	5.1	504	1	GUNW_ERWCA
11	88.5	5.1	509	1	GATB_MYCLE
12	87.5	5.1	1396	1	VLTF_BPT5
13	87	5.0	573	1	Y2G8_MYCLE
14	86.5	5.0	454	1	RHSE_RHIME
15	86.5	5.0	554	1	IJUK_ECOLI
16	85.5	5.0	474	1	UCR2_FUGGR
17	85.5	5.0	509	1	GATB_MYCTU
18	85.5	5.0	1275	1	RFBC_MVXXA
19	85	4.9	456	1	REP_STRLI
20	84.5	4.9	365	1	EL3B_NICPL
21	84.5	4.9	370	1	EL3B_TOBAC
22	84.5	4.9	371	1	EL3B_TOBAC
23	83	4.8	416	1	PROA_STRTR
24	83	4.8	573	1	Y2G8_MYCTU
25	82.5	4.8	572	1	SYM_AERPE
26	82.5	4.8	1278	1	DHBE_BACSU
27	82	4.8	429	1	RSI_LEULA
28	81.5	4.7	383	1	MEXA_PSEAE
29	81.5	4.7	544	1	CH61_SYNVU
30	81.5	4.7	1234	1	MF0_MYCTU
31	81	4.7	154	1	YD88_SYNY3
32	81	4.7	451	1	Y0D0_MYCTU
33	81	4.7	559	1	ILVB_KLEPN

34 81 4.7 918 1 IMB2_YEAST
35 80.5 4.7 532 1 IPA7_SHIFL
36 80 4.6 541 1 CH60_CYAPA
37 80 4.6 810 1 CLPC_BACSU
38 79.5 4.6 339 1 RFAL_ECOLI
39 79.5 4.6 574 1 IPA4_SHIFL
40 79.5 4.6 718 1 CTPC_MYCTU
41 79.5 4.6 1829 1 FRPC_NEIME
42 79 4.6 547 1 SAPA_ECOLI
43 79 4.6 657 1 CTPD_MYCTU
44 79 4.6 759 1 YEHM_ECOLI
45 79 4.6 947 1 SECA_STRLI

ALIGNMENTS

RESULT 1
DMPL_PSESP
ID DMPL_PSESP STANDARD; PRT; 331 AA.
AC P19730;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PHENOL HYDROXYLASE P1 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE
DE P1 COMPONENT).
GN DMPL OR PHEA2.
OS Pseudomonas sp. (strain CF600).
OG Plasmid pVII50.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91072230.
RA Nordlund I., Powlowski J., Shingler V.;
RT *Complete nucleotide sequence and polypeptide analysis of
RT multicomponent phenol hydroxylase from Pseudomonas sp. strain
RT CF600.*;
RL J. Bacteriol. 172:6826-6833(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BH;
RA Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,
RA Fujita M.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES. P1 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR
CC IN VITRO PHENOL HYDROXYLASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD, AND REQUIRES FE(2) FOR ACTIVITY.
CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED
CC BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M60276; AAA25940.1; -;
DR EMBL: D28864; BAA06015.1; -;
DR PIR: B37831; B37831.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW NADP; Flavoprotein; FAD; Iron; Plasmid.
SQ SEQUENCE 331 AA; 38208 MW; 8042F5723BE3A5E8 CRC64;

Query Match 45.2%; Score 779.5; DB 1; Length 331;
Best local similarity 46.8%; Pred. No. 5.3e-60;
Matches 155; Conservative 54; Mismatches 115; Indels 7; Gaps 4;


```

Query Match          7.1%; Score 123; DB 1; Length 394;
Best Local Similarity 22.3%; Pred. No. 0.0022;
Matches 74; Conservative 49; Mismatches 167; Indels 42; Gaps 14;

QY 28 KTAIRYQECGMGAQPO-----ENFH-YRPTWDPDYEIFDPSAISARMANWY 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 47 KPUSEYEQISCYAPNDPDIAGLDGWDWTQKPHGGPSWGN-----STELRTDWW 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 ALKDP-RQFY--YASWATTRARQODAMESFEFVESRMITG-LMRDDVAARALDVLPLR 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 RHEDPARWHHPYVKDSEARYIQRLAAYSSEGSTETDTPYRDEHLILNKYFGALYSE 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 HAAWGANNMNAQICALGYGVTFAPAMFAMDNLGVQO--YLTELALAMAEPPDV---LEA 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 YGLFNAHSSVRDC--LSDTIQTAFFAALDKVDNAQIMQMERLFIATKLVPGFDASTDV 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 AKATWTFDAAWQPLRRYVDETIV-VADPVELFIAQNLALDGLLYPLVYDRFVDERIALEG 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 PKIWTWTDPIYSGARATVOEIVGQVDNTEILWAGHAVMTATFGQFARREFFORLATIYIG 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 GSAVAMLTAPMPWHETNESWIDAV-VKTMAESD---DNRALLARWTQDSARAEEALA 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 DTILTFPTAQSYQFQTRGAIDDLFVYCLANDESEFGAHNRTFLNATHEYLASSVAALK 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 299 P-YAARALQDAGRAALDEVREQPHARAARLGI 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 DFVGLYAKVEKSR--DRSRRLRGAAASSAI 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
GUNN_ERWCA STANDARD; PRT; 444 AA.
AC ID GUNN_ERWCA STANDARD; PRT; 444 AA.
AD Q59394;
DT DT 01-NOV-1997 (Rel. 35, Created)
DT DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ENDOGLUCANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE N
  (CELLULOSE N).
DE DE DE
GN GN
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATROSEPTICA FCBR C18;
RX MEDLINE; 98299944.
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
  von Wettstein D.;
RA "Transplanting two unique beta-glucanase catalytic activities into
  one multienzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
  LINKAGES IN CELLULOSE..
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
  HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation in
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed, usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/).
  or send an email to license@isb-sib.ch.
CC -----
CC EMBL; L39788; AAC37033.1; -.
CC DR PFAM; PF00942; CBD_3; 1.
CC DR PFAM; PF00150; cellulase; 1.
CC DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31
FT CHAIN 32 444
FT ENDOGLUCANASE N.

```


RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC -!- SIMILARITY: TO M.TUBERCULOSIS RV1563C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z74020; CAA98329.1; -
 DR TUBERCULIST; RV1562C; -
 DR PRAM; PF00128; alpha-amylase; 1.
 KW Hypothetical protein; Hydrolase; Glycosidase.
 SQ SEQUENCE 580 AA; 64076 MW; 506190468F4B862 CRC64;

Query Match 5.7%; Score 99; DB 1; Length 580;
 Best Local Similarity 19.8%; Pred. No. 0.43; Indels 126; Gaps 14;
 Matches 67; Conservative 37; Mismatches 108; Indels 126; Gaps 14;
 QY 54 PDYEIFDSRAIRN-----ANWY-----ALKDPRQ 79
 Db 2 PEFRWAFKPAVLVDVNGAVHAMTRSDGWHTTVAADARYGLLDPTVLPDPRS 61
 QY 80 FYASWATTRAQDAMESNEFEVSRMIGLMRDDVAARALDVLPLRHAW-GANNNN 138
 Db 62 -----ARQPDGVHARSQRPEPPQFGAARTD-----TCWFGRSVEG 97
 QY 139 AQICALGYGTFTAPAMFAMDNLGVAOYLRLALAMAEPDVLEAAKATWTRDAWQPIR 198
 Db 98 AVIVELHIGTFTAGTTPDAIEKL---DYLVDLGIDFVE-----ER 237
 QY 199 RYVEDTLVVADEVLELFT-AQNIALDGLLYPLVYD-----RFVD----- 237
 Db 134 -----LMPVNSFAGTRGNGYDGLWVSYHEPYGGPDGIVRFIDACHARRLGLVLDA 184
 QY 238 IALEGGSVAMLTAFMPHWHHTESNRWIDAVVKTMAESDDNFALL-----ARKWTRWSA-- 291
 Db 185 VFNHLGPSNGYLPREFGPYISSASNPWSDG-INIAGADSDEVRHYIIDCALRWMRDFHADG 243
 QY 292 -RAEALAPVAARALQDAGRALDREVRQEHARAARLG 328
 Db 244 LRLDAVHALVDTTAVH-----VLELANATRWLSGQLG 276

RESULT 8
 Y568_MYCTU STANDARD; PRT; 472 AA.
 AC OS3765;
 DT 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DE PUTATIVE CYTOCHROME P450 RV0568 (EC 1.14.-.-).
 GN RV0568 OR MT039.06.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE; 98295987.
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AL021942; CAA17439.1; -
 DR TUBERCULIST; RV0568; -
 DR PRAM; PF00067; P450; 2.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 388 HEME (BY SIMILARITY).
 SQ SEQUENCE 472 AA; 50687 MW; B7F2898BEE80863F CRC64;

Query Match 5.3%; Score 91.5; DB 1; Length 472;
 Best Local Similarity 25.2%; Pred. No. 1.5;
 Matches 41; Conservative 20; Mismatches 69; Indels 33; Gaps 4;
 QY 162 LGVAOYLRLALAMAEPDVLEAAKATWTRDAWQPIRKYVEDTLVVAOPVELFIAQNLA 221
 Db 177 LNVGWPAT---LATANPSLL-----NNLWLSRLRIIEEA----- 208
 QY 222 DGLLYPLVYDREVDIERIALEGGSVAMLTAFMPHWHHTESNRWIDAVVKTMAESDDNRA- 280
 Db 209 DALLVAETADRADPDLAARTDTLAMLVRAEDGEGTWTETRELROQLITLLVAGHTTAT 268
 QY 281 -----LLAKWTRWSARAFAALAPVAARALQDAGRALDREVRQ 319
 Db 269 GLSWALERLTHRPVTLAKAVQAADASAAGDPAGDEYLDVAKE 311

RESULT 9
 KLB2_ECOLI STANDARD; PRT; 461 AA.
 ID KLB2_ECOLI
 AC P52605;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE KLCB PROTEIN.
 GN KLCB.
 OS Escherichia coli.
 OG Plasmid Incp-alpha RK2.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94285211.
 RA Pansegrau W., Lanka E., Barth P.T., Figurski D.H., Guiney D.G.,
 RA Haas D., Helinski D.R., Schwab H., Stanisich V.A., Thomas C.M.;
 RT "Complete nucleotide sequence of Birmingham Incp alpha plasmids.
 Compilation and comparative analysis.";
 RL J. Mol. Biol. 239:623-663(1994).
 RN [2]
 RP SEQUENCE OF 1-237 FROM N.A.
 RA Larsen M.H., Figurski D.H.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- DOMAIN: CONTAINS A DNA-BINDING REGION JOINED BY A SHORT VARIABLE
 CC SEGMENT TO A REGION SIMILAR TO E-COLI KORA AND TRBA.

CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L27758; NOT ANNOTATED_CDS.
 DR EMBL: U05773; AAA57449.1; -
 KW Plasmid; DNA-binding; Transcription regulation.
 FT SIMILAR 315 344 TO THE C-TERMINAL OF E.COLI KORA AND
 FT TRBA.
 SQ SEQUENCE 461 AA; 51165 MW; 10EBFB2F2E00BD81 CRC64;

Query Match 5.1%; Score 88.5; DB 1; Length 461;
 Best Local Similarity 22.5%; Pred. No. 2.6;
 Matches 56; Conservative 31; Mismatches 73; Indels 89; Gaps 14;

QY 90 AROQDAMESNEFFVESRMIGMRDDVAARALDVLPLRHAAG-----AMNNAQI-- 141
 DB 171 AGQDAAE-----RVI-----ERHCRATPGVPM-----WGQSGDFLASVDGMRYWV 212
 QY 142 -CALGYGTGTAPAMFAMD-----NLGVAQYLTRLALAMAEPDVLEAAKATWTRDAA 193
 DB 213 EVESYGGITLVHPQFHAVLDLDPFFISFTGYRSHYDHARGMTVDQVADGVLRALLSH- 271
 QY 194 WQPLRVVDETLVADPVELFIAQNLADGLLYPLVDFRVDERIALEGGSVAMLT--- 250
 DB 272 -----RRYID-----ARD-----QDLADEPLP-----AWLAGITPPP 299
 QY 251 -----AFMPENHETSN-----RWIDAVYKTMAESDDNRALLARWTRDWSARAALAPVA 301
 DB 300 RRVAVVEDMKPDELPGPAWVDVLP-----HQAFTIA-----RKWAASAKAKLAAR 349
 QY 302 ARALQDAGR 310
 DB 350 AKAQEPAGQ 358

RESULT 10
 GUNW_ERWCA STANDARD; PRT; 504 AA.
 AC Q59395;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ENDOGLUCANASE VI PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V1)
 DE (CELLULOSE V1).
 GN CELVI.
 OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwinia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SC3193;
 RX MEDLINE: 95231512.
 RA Mae A., Heikinheimo R., Palva E.T.;
 RT "Structure and regulation of the Erwinia carotovora subspecies
 RT carotovora SC3193 cellulase gene celVI and the role of cellulase in
 RT phytopathogenicity.";
 RL Mol. Gen. Genet. 247:17-26(1995).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X79241; CAA55823.1; -
 DR HSSP: Q06851; LNBC.
 DR PFAM: PF00942; CBD_3; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 504 ENDOGLUCANASE V1.
 FT DOMAIN 32 334 CATALYTIC.
 FT DOMAIN 335 352 LINKER.
 FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 5.1%; Score 88.5; DB 1; Length 504;
 Best Local Similarity 23.4%; Pred. No. 2.9;
 Matches 65; Conservative 31; Mismatches 99; Indels 83; Gaps 16;

QY 112 MRDDVAARALDVLPLRHAAGCAN-----MNAQICALGYGTGTAPAMFAMDNLGVA 165
 DB 81 LRDDGIVNFRVAMTYAENGIANSLANKYKVAAGGLGYIID--MHTLSDNPN 138
 QY 166 QYLTRLALAMAE-----PDVL-EAA-----KATWTRDAWQPLRRYVEDTLVADPV 211
 DB 139 TYKAAKIFFEAMAGLYGNSPNVIYEIANEPNGSVTW--NGQIRPYALEVDTITRSKDPD 196
 QY 212 ELFI-----AQNIALD-GLLYPL-----VYDFVDERI--ALEGGSVAMLT 250
 DB 197 NLIIVSGTWSQDTHDAADNOLPDPTLYALHFYAGTHQGLRDRIDYQAQSGAAI---- 252
 QY 251 AFMPWHNT-----ESNRWIDAVKTMAESDDNRAL-LARWTRDWSARAAL 297
 DB 253 -FVSEWGTSDASNGGPPPLPSQWIDFL-----NNRGISWVNSISDKSSTAAL 302
 QY 298 APVAARA-----IQDAGRAALDEVREQFHARAALG 328
 DB 303 VAGASKSGGWTEQNLTSTGKF---VREQIRAGAGLSG 336

RESULT 11
 GATB_MYCLE STANDARD; PRT; 509 AA.
 AC Q33107;
 DT 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
 DE SUBUNIT B).
 GN GATB OR MLCB637.15.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Oliver K., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases
 CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
 CC GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-
 CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
 CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
 CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
 CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
 CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GATB FAMILY.

Query Match 5.0%; Score 87; DB 1; Length 573;
 Best Local Similarity 22.8%; Pred. No. 4.5;
 Matches 82; Conservative 37; Mismatches 123; Indels 118; Gaps 17;

QY 25 GCDKATRYQEGMGAQPOENF-----HYRPT-----WDPYEIFDPSRS 64
 DB 220 GNEDAVALLEWLTTHPESKYSAAKDPVRLTTTAAEQIAARADPWPDPSSVWTDNSDR 279
 QY 65 AIRMANWVALKDPPOFYFASWATTRARQODAMESFEVESRMIGLMDVDAARALDVL 124
 DB 280 DRLTLQAQALD-RQI-----GLTRVKIQ-----IERYAATMMKVAAGMKVA 324
 QY 125 VPLRHAANGANNNAQICALGYGTFTAPAMPAMDNLGVAQYTLRLALAMAEPDVLAA 184
 DB 325 OPSKHMIEGT-----PPGFGKTTIARVVANILAGLV-----ISEPKLVETS 366
 QY 185 KATWTRDAWQPLRYVEDTLVADPVELFTAQNALDGLL-----YPLVDRFVDERIA 239
 DB 367 RDRFVAETEGO-----SAVKAAKTIDIALGVLFDIDAYALVOER--DGRTD 411
 QY 240 LEGGSVAAMLTAPEWHTESNRWIDAVVKTMAESDD-----NRALLARW-TR--- 287
 DB 412 PFGQEAULTLARM-----ENDR--DRLVLIAGYSSDIDRLLETNEGRSFATRIEFD 464
 QY 288 -----DWSARAE-----ALAPVAARALODAGRAALDEVREQFHAR 323
 DB 465 TYPPELLEIAKVIATDADSLSAEASKNLLEAKQLAQLTLR--GRPALDVAGNGRYAR 522

RESULT 14
 RHSE_RHME STANDARD; PRT; 454 AA.
 AC Q9Z3Q8;
 DT 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE RHIZOACTIN SIDEROPHORE BIOSYNTHESIS PROTEIN RHSE (EC 1.-.-.-).
 GN RHSE.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymA (megaplasmid 1).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RCR2011 / S047;
 RA Lynch D., O'Connell M., O'Brien J.;
 RT "Cloning and sequence analysis of the Rhizobium meliloti 2011
 rhizobactin regulon."
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- PATHWAY: RHIZOACTIN SIDEROPHORE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE LYSINE N6-HYDROXYLASE / L-ORNITHINE N5-
 OXYGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF110737; AAD09416.1; -
 KW Oxidoreductase; Iron transport; NADP; Flavoprotein; FAD; Plasmid.
 FT NP_BIND 7 13 FAD (POTENTIAL).
 SQ SEQUENCE 454 AA; 50211 MW; 57074278FFAFCID1 CRC64;

Query Match 5.0%; Score 86.5; DB 1; Length 454;
 Best Local Similarity 23.0%; Pred. No. 3.7;
 Matches 66; Conservative 40; Mismatches 102; Indels 79; Gaps 16;

QY 28 KTATRYOEGM--GAQPOENF--HYRPTWDPYEIFDPSRSAL--RMANNYALKD---PR 78
 DB 36 KPAFRHEGLIIPGTLQVFPFADLVMTADPTRLUSFLNVLAVHDLRLKYFYFENPMIR 95
 QY 79 QFY--YASWATTRARQODAMESFEVESRMIGLMDVDAARAL--DVLVPLRHAAGA 134
 DB 96 QEYDHYCRWA---SQQLSACRFGEEV-----DVAHESADSFLVESRSASGSK 141
 QY 135 NNNNAQICALGYGT-----VFTAPAMFAMD-----NLGVAQYLTR 170
 DB 142 QOYRSNTAIGVGTAFPLPKWAQIKTLAPLHMSSEFGRRLSKRRRVTVIGSGSAAE 201
 QY 171 LALANAE---PDVLEA-AKATW--TRDAWQ-----LRRYVEDTL----- 205
 DB 202 CVIALNLDTPEWVAAGASIQWITRSAGEFPMEYSKLGLLEYTPDVMRHFHRIAPVRRRE 261
 QY 206 VVADVELFIAQNALDGLLYPLVDRFY---DERIALEGSSAVML 249
 DB 262 IVADQGLLYKGISFTIGTIGFDLMYERSVGGRRDPGLALFNSCAVETL 308

RESULT 15
 YJJK_ECOLI STANDARD; PRT; 554 AA.
 AC P37797;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE ABC TRANSPORTER ATP-BINDING PROTEIN YJJK.
 GN YJJK.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE; 95334362.
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes."
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [2]
 RP REVISIONS TO C-TERMINUS.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE; 97426617.
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE OF 1-79 FROM N.A.
 RX MEDLINE; 92041559.
 RA Engel H., Kazemier B., Keck W.;
 RT "Murein-metabolizing enzymes from Escherichia coli: sequence analysis
 RT and controlled overexpression of the slt gene, which encodes the
 RT soluble lytic transglycosylase."
 RL J. Bacteriol. 173:6773-6782(1991).
 RN [4]
 RP PARTIAL SEQUENCE OF 1-11.
 RC STRAIN=K12 / EMG2;
 RX MEDLINE; 97443975.
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12."
 RL Electrophoresis 18:1259-1313(1997).
 RN [5]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX MEDLINE; 99420866.
 RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
 RT "Enrichment of low abundance proteins of Escherichia coli by

```
RT hydroxyapatite chromatography." ;
RL Electrophoresis 20:2181-2195(1999).
CC -|- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). EF-3 SUBFAMILY. STRONG, TO H.INFLUENZAE
CC H11252.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14003; AAA97287.1; ALT_FRAME.
DR EMBL; AE000509; AAC77344.1; -.
DR EMBL; M69185; -; NOT_ANNOTATED_CDS.
DR ECOGENE; EG12343; YUJK.
DR FFAM; PF00005; ABC_tran; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport; Repeat.
FT INIT_MET 0
FT NP_BIND 38 45 ATP (POTENTIAL).
FT NP_BIND 355 362 ATP (POTENTIAL).
SQ SEQUENCE 554 AA; 62311 MW; 444C605C67A019B1 CRC64;
```

Query Match 5.0%; Score 86.5; DB 1; Length 554;
Best Local Similarity 17.8%; Pred No. 4.8; Mismatches 129; Gaps 14;
Matches 67; Conservative 50;

QY	10	IKPLRHTFAHVA-----ONIGGDKTATRYQEGM-----MGAQ 41
Db	14	VPPKRHIILKNISLSPFGAKIGVLGLNGAGKSTLLRIMAGIDKIDIEGEARPOPDIKIGYL 73
QY	42	PQENFHYRTWPDVEIFPDSRAI-----RMANWTALKDPROFYASWATTTARQQ 93
Db	74	POE-----PQLNPEHTVRESIEEAVSEVVNALKRDLDEVYALYADPDADFCLKAAEQGRLE 128
QY	94	DAMESNFEVESRRMIGLRDVAARALDVLVPLRHAAGANNNAQICAIQGYGTVFTAP 153
Db	129	ETIQAH-----DGHMLNVQLERAADA---LRLPDWDAKIANTL----- 162
QY	154	AMFHAMDNLGVAQYLTRLALAMAEPDVLEAAKATWTRDA---AWQPLRRYVEDTLVYADP 210
Db	163	-----SGGERRRVALCRLLLEKPDMLLDDEPTNHLDAESVAV--LERFLHD----- 206
QY	211	VELFTAQNLALDGLLYPLVYDRFVDERIA-----LEGGSVAVMLTAFMPFWHTESNRWID 265
Db	207	-----FEGTVVAITHDRYFLDNVAGWILELDRGEGI-----PWEGNYSWLE 248
QY	266	AVVKTMAESDDNRLARWTRDWSARAAPALAPVAARALQDAGRAAL----- 313
Db	249	OKDQRLAQASQEA-----RRKSIEKELEWVRQTKGRQSKGARLARFAELNSTEYQ 302
QY	314	--DEVREQFHARAARLG 328
Db	303	KRNETNELFIPPGPLRG 319

Search completed: September 26, 2000, 20:30:02
Job time: 426 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:11:41 ; Search time 75.19 Seconds
(without alignments)
28.036 Million cell updates/sec

Title: US-09-430-029-4

Perfect score: 451
Sequence: 1 MSNVVFAQANESRPVDA.....ITLSGHIDEDEFTLSWSH 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	355	78.7	92	1 W06801	Toluene ortho-mono
2	257	57.0	90	1 W98972	Alcaligenes sp. pr
3	67.5	15.0	410	1 R51368	Protein containing
4	65	14.4	1093	1 R40815	Truncated IBV spik
5	65	14.4	1162	1 P70137	Sequence of a regi
6	65	14.4	1162	1 W06829	Major neutralising
7	64.5	14.3	394	1 W97744	Wheat S-adenosylme
8	64	14.2	292	1 W80988	Amino acid sequenc
9	64	14.2	1162	1 P60720	Sequence of the SI
10	63.5	14.1	394	1 W34540	S-adenosylmethioni
11	63.5	14.1	394	1 R34541	S-adenosylmethioni
12	63.5	14.1	394	1 W34542	S-adenosylmethioni
13	62.5	13.9	406	1 R11349	Cytochrome enzyme
14	62	13.7	322	1 R61118	Yeast presequence
15	62	13.7	897	1 R20982	Sequence of beta-c
16	61	13.5	256	1 W01203	Serine protease pf
17	60.5	13.4	333	1 R39490	Human apoAIV mutel
18	60.5	13.4	342	1 R39491	Human apoAIV mutel
19	60.5	13.4	1050	1 W67642	A serine/threonine
20	60.5	13.4	1168	1 R16111	Spike protein of i
21	60.5	13.4	2639	1 R73476	Grapevine leafroll
22	59.5	13.2	339	1 R89748	AFT-1 interacting
23	59.5	13.2	464	1 W25116	CD2-associated int
24	59.5	13.2	464	1 W26496	CD2 associated int
25	59.5	13.2	464	1 W80420	CD2 associated int
26	59.5	13.2	553	1 W25115	CD2-associated int
27	59.5	13.2	553	1 W26495	CD2 associated int
28	59.5	13.2	553	1 W80419	CD2 associated int
29	59.5	13.2	659	1 W37724	CD2 associated int
30	58.5	13.0	732	1 R11056	Acylamino acid-iso
31	58	12.9	117	1 R6214	Protein encoded by
32	58	12.9	117	1 R81470	Nocardia corallina
33	58	12.9	681	1 R72970	GLVR-1 protein. Ne

34 58 12.9 681 1 W96999 The gibbon ape leu
35 58 12.9 1477 1 R67691 S. cerevisiae scau
36 58 12.9 1477 1 W10424 Saccharomyces cere
37 58 12.9 1477 1 Y08819 Fumonisin-resistan
38 57.5 12.7 504 1 W36140 Bovine p58 protein
39 57 12.6 1051 1 W67641 A serine/threonine
40 56.5 12.5 297 1 Y07078 Renal cancer assoc
41 56.5 12.5 355 1 W29099 Polyamide hydrolas
42 56.5 12.5 355 1 W36011 Polyamide hydrolas
43 56.5 12.5 618 1 R94587 Dnak protein. Dnak
44 56 12.4 459 1 Y11069 H. pylori ORF 12ae
45 56 12.4 559 1 W19630 Streptomyces venez

ALIGNMENTS

RESULT 1

W06801

ID W06801 standard; Protein; 92 AA.

AC W06801;

DT 29-JAN-1997 (first entry)

DE Toluene ortho-monoxygenase subunit tomA2.

KW tom; PTOM; self-transmissible; constitutive; bioreactor; pollutant;

KW breakdown; trichloroethylene; TCE; degradation.

OS Pseudomonas cepacia strain PRL-23.

PN US5543317-A.

PD 06-AUG-1996.

PF 02-MAY-1991; 694718.

PR 02-MAY-1991; US-694718.

PR 15-DEC-1993; US-167457.

PR 06-OCT-1994; US-319387.

PA (FRAN/) FRANCESCONI S C.

PA (SHIE/) SHIELDS M S.

PI Francesconi SC; Shields MS;

DR WPI; 96-370640/37.

DR N-PSDB; T44457.

PT Microorganisms transformed with P. cepacia PRL-23 Tom enzyme gene -
Claim 1: Column 29-30; 25pp; English.

PS The present sequence is that of toluene ortho-monoxygenase subunit

CC tomA2 encoded by T44457, isolated from Pseudomonas cepacia strain

CC PRL-23. The Tom gene is present on a large self-transmissible plasmid

CC denoted PTOM. The enzyme is capable of degrading trichloroethylene

CC (TCE), a hazardous pollutant. The PTOM plasmid is transmissible and

CC expressible in other bacteria, thus many bacteria can be genetically

CC altered to constitutively degrade TCE, esp. in bioreactors or

CC TCE-contaminated environments. P. cepacia PRL-23 contg. PTOM does not

CC need exogenous chemical inducers and is capable of functioning under a

CC diverse set of conditions. Also it does not require an inducer that is a

CC co-substrate for the enzyme required to break down TCE, in effect, it is

CC not subject to competitive inhibition.

CC Sequence 92 AA;

SQ

Query Match

Best Local Similarity 78.7%; Score 355; DB 1; Length 92;

Matches 65; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

OY 1 MSNVVFAQANESRPVDAIVADNPRVAVVSPGMVKIDAPDLATRTTIEELTGRF 60

DB 3 MSNVVFAQANESRPVVEAILADNPKAVATESPGMVKIDAPGHLTINRQSIDRIGMKF 62

OY 61 DLQQLVNLITLSCHIDEDEFTLSWSH 89

DB 63 DLQQLVNLITLSCHIDEDEFTLSWSH 91

RESULT 2

W98972

ID W98972 standard; Protein; 90 AA.

AC W98972;

DT 10-MAY-1999 (first entry)

DE Alcaligenes sp. protein PoxC.
KW Alcaligenes; PoxR; PoxA; PoxB; PoxC; PoxD; PoxE; PoxF; PoxG; PoxH;
KW PoxI; aromatic; oxidative; petroleum purification; chemical industry;
KW drug.
KW OS Alcaligenes sp.
PN J11042088-A.
PD 16-FEB-1999.
PF 25-JUL-1997; 200625.
PR 25-JUL-1997; JP-200625.
PR (TOFU) TONEN CORP.
PA WPI; 99-197820/17.
DR N-PSDB; X18867.
DR N-PSDB; X18867.
PT New encoding an aromatic cpd. oxidative decompn. enzyme - useful in
PT the fields of petroleum purification, chemical and drug industries
PS Claim 1; Page 7-19; 3pp; Japanese.
CC The present sequence encodes Alcaligenes sp. protein PoxC. The present
CC invention also describes PoxA, PoxB, PoxR, PoxD, PoxE, PoxF, PoxG, PoxH,
CC and PoxI. The proteins are useful in the fields of petroleum
CC purification, chemical industry and drug industry related to the
CC synthesis, conversion and decomposition of aromatic compounds.
SQ Sequence 90 AA;

Query Match 57.0%; Score 257; DB 1; Length 90;
Best Local Similarity 57.5%; Pred. No. 7e-26;
Matches 50; Conservative 16; Mismatches 21; Indels 0; Gaps 0;

Qy 2 SNVYIAFOANEDSRPIVDVAIVADNPRAVVESPGMWKIDAPDLRTTRETIELTGTRFD 61
:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 3 ANYVIALQNDDTRPIIEAIAEANPLAVVSQFFAMWKIDAPGRLTTVRELVAOKLGRWD 62

```
QY 62 LQQLQVNLITLSCHIDEDEFTLSWS 88
    ||:::||:|||||:||||| 89
Db 63 LQEIHLNLSLSCNIDETDEAFTLHWS 89
```

RESULT	3	
R51368		
ID	R51368 standard; Protein; 410 AA.	
AC	R51368;	
DT	24-NOV-1994 (first entry)	
DE	Protein containing Cytochrome P450 SCA-2 activity.	
KW	Cytochrome P450 SCA-2; Streptomyces carbophilus;	
KW	treatment of hyperlipidaemia; drug preparation.	
OS	Streptomyces carbophilus.	
OS	Streptomyces carbophilus.	
FH	Location/Qualifiers	
FH	1. .410	
FT	protein	
FT	/label= cytochrome P-450 containing activity	
PD	J060707080-A.	
PD	15-MAR-1994.	
PF	28-AUG-1992; 229969.	
PR	28-AUG-1992; JP-229969.	
PA	(SANY) SANKYO CO LTD.	
PA	WPI; 94-128679/16.	
DR	N-PSDB; Q61452.	
PT	Cytochrome P450 SCA-2 gene - from Streptomyces carbophilus,	
PT	useful in treatment of hyperlipidaemia	
PS	Claim 1; Page 12-14; 18pp; Japanese.	
CC	R51368 shows a protein having cytochrome P-450 activity. P-450	
CC	SCA-2 can be prepared commercially for use in the preparation of a	
CC	drug for the treatment of hyperlipidaemia.	
SC	Sequence 410 AA.	

Query Match 15.0%; Score 67.5; DB 1; Length 410;
Best Local Similarity 31.2%; Pred. No. 1.1;
Matches 24; Conservative 11; Mismatches 25; Indels 17; Gaps 3;

QY	13	DSRPIVDAIVADNP-----RAVVVSPGMKIDAPDLRTIRRETIELT-----GTRED	61
Db	70	DPLRSSDLRHADFATSPRFKAFFQSPAFIGNDPPEHGTTRRRMTISEFTVKRKGMRPD	129
QY	62	LOOLQVNLITLSGHIDE	78

```

Db      130 VERI-----VHGFIDD 140
          : : : : : | | | :

```

RESULT	4	
R40815		R40815 standard; Protein; 1093 AA.
ID	AC	R40815;
DT	11-FEB-1994	(first entry)
DE	Truncated IBV spike protein.	
KW	Infectious bronchitis virus; IBV; spike protein; transmembrane;	
KW	cytoplasmic; anchor; vaccine; poultry; fowl.	
OS	Infectious bronchitis virus strain M41.	
PN	GB2264497-A.	
PD	01-SEP-1993.	
PF	17-FEB-1993;	003123.
PR	19-FEB-1992;	GB-003509.
PR	(BTE-) BRITISH TECHNOLOGY GROUP LTD.	
PI	Skinner MA;	
DR	WPI; 93-275023/35.	
DR	N-PSDB; Q47743.	
PT	DNA encoding truncated spike protein of infectious bronchitis	
PT	virus - used to express protein to protect fowl, esp. poultry,	
PT	against viral infection	
PS	Claim 1; Page 15-23; 43pp; English.	
CC	The gene (Q47743) encodes a spike protein (R40815) which is located	
CC	the transmembrane and cytoplasmic anchor regions of native IBV.	
CC	The truncated spike protein can be used to vaccinate fowl, esp	
CC	from IBV infection.	
CC	Sequence 1093 AA;	
SQ		SQ

Query Match 14.4%; Score 65; DB 1; Length 1093;
Best Local Similarity 26.7%; Pred. No. 8.6;
Matches 28; Conservative 11; Mismatches 46; Indels

```

4  VFTAFQANEDSRFIDAIVA-----DNPRVVVESGVMKIDAPDLRTTR 48
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
934 VFTSYFTSDSFYNVAIGFCVKPANASQAIAPVANGRGIFTQVNGSYIATDMYMR 993
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

49 RETIEEL-TGTRFDLQLOLVNLTLSCHIDEED-----DEFTLSWS 88
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
994 AITAGDIVILTSCQANKSVNKIVITTFVNDDEDFENDELSKWN 1038
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 5
P70137
ID P70137 standard; protein: 1162 AA.

AC	P/0137;
AD	09-APR-1991 (first entry)
AE	Sequence of a region of the S1 protein of infectious
DE	bronchitis virus (IBV) strain M41.
DE	Poultry vaccine; antigen; immunogen.
KW	Infectious bronchitis virus.
OS	EP-221609-A.
PN	EP-221609-A.
PD	13-MAY-1987.
PD	28-OCT-1986; 201884.
PF	31-OCT-1985; NL-002975.
PR	(DUIN) DUPHAR INT RES BV.
PR	PI Nieters HCM, Van der Zeijst HAM; .
PI	WPI; 87-130709/19.
DR	N-PSDB; N70196.

N-9526; N70190.
 The gene coding for protein of infectious bronchitis viruses -
 useful in prodn. of peptide or protein for vaccination of poultry
 against the viruses.
 Claim 2; Fig 1; 21pp: English.
 Proteins encoded by IBV strains M41, D1466, D207 or D274are claimed
 together with the genes that encode them, and the gene that encodes
 a protein from IBV strain H120. In addition, a peptide is claimed,
 characterised in that it comprises at least one of the regions in
 which the AA sequences of the IBV strains M41 and M42 are
 different, esp. the region between AAs 55 and 70 and/or between AAs
 116 and 131 and/or between AAs 230 and 301.

DE Amino acid sequence of the human TSC-22-like protein-2.
 KW Human; TSC-22-like protein-2; HT22L-2; respiratory disorder;
 KW neurological disorder; antibody.
 OS Homo sapiens.
 PN W09850425-A2.
 PD 12-NOV-1998.

PF 04-MAY-1998; U09122.
 PR 05-MAY-1997; US-851190.
 PA (INCY-) INCYTE PHARM INC.
 PI Goli SK, Hillman JL;
 DR WPI: 99-034708/03.
 DR N-PSDB; V68643.

PT New human TSC-22-like protein - for treatment, prevention and
 PT diagnosis of developmental, respiratory and neurological disorders
 PS Claim 18; Fig 2: 65pp; English.
 CC This is the amino acid sequence of the human TSC-22-like protein-2
 CC (HT22L-2) used in the method of the invention used in the treatment,
 CC and diagnosis of developmental respiratory and neurological disorders.
 CC Other uses for HT22L-2 include generation of antibodies and screening
 CC for specific binding agents (potential therapeutics). HT22L-2 and its
 CC fragments are used as antisense/ribozyme therapeutics, for detecting
 CC and quantifying gene expression (as probes and primers in standard
 CC hybridisation and amplification assays), for isolating related
 CC sequences and for chromosome mapping. Antibodies are used directly
 CC as therapeutic antagonists, as immunoassay reagents for diagnosis or
 CC monitoring disease or treatment, in competitive drug screens and to
 CC isolate HT22L-2 from natural sources.
 SQ Sequence 292 AA;

Query Match 14.2%; Score 64; DB 1; Length 292;
 Best Local Similarity 24.6%; Pred. No. 2;
 Matches 17; Conservative 13; Mismatches 21; Indels 18; Gaps 2;

QY 13 DSRPIVDAIVNPRVAVVSPGMYKIDAPDLRTIRRIETIELTGTRFDLQLOVNLITL 72
 DB 20 DSRPSPALYFTHDSLHKSP-----DFFGVAQAQKFSLAH---SMLAI 61
 QY 73 SGHIDEDD 81
 DB 62 SGHLSDDD 70

RESULT 9
 ID P60720 standard; Protein; 1162 AA.
 AC P60720.
 DT 22-JUL-1991 (first entry)
 DE Sequence of the S1 signal and S1 and S2 polypeptides of the spike
 DE protein.
 KW Poultry; infectious disease; diagnosis; probe; virus-typing.
 OS Infectious bronchitis virus (IBV) Beaudette strain.
 FH Key Location/Qualifiers
 FT peptide 1..18
 FT region /label= signal
 FT 19..537
 FT region /label= S1
 FT 538..1162
 FT region /label= S2

PN W08605806-A.
 PD 09-OCT-1986.
 PF 27-MAR-1986; G00181.
 PR 29-MAR-1985; GB-008265.
 PR 20-JUN-1985; GB-015678.
 PA (NATR) NATIONAL RES DEV CORP.
 PA (BINN/) BINNS M M.
 PI Binns MW, Boursnell MEG, Brown TDK;
 DR WPI: 86-278819/42.
 DR N-PSDB; N60679.
 PT Artificial infectious bronchitis virus spike protein - and
 PT corresp. DNA molecules useful as probes for the infection
 PS Disclosure; Page 6-11; 44pp; English.
 CC The patentors claim a sequence which codes for (1) the spike protein

CC precursor; (2) the S1 signal plus the S1 polypeptide; (3) the S1
 CC polypeptide; or (4) the S1 polypeptide plus the S2 polypeptide, pref.
 CC of IBV Beaudette M41 or 6/82 strain.
 SQ Sequence 1162 AA;

Query Match 14.2%; Score 64; DB 1; Length 1162;
 Best Local Similarity 26.7%; Pred. No. 13;
 Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

QY 4 VFIAFQANEDSRPIVDAIVA-----DNPRAVVVSPGMYKIDAPDLRTIR 48
 DB 934 VFHFISYTPDSFVNVTAIYVCFKPNASQYAIYVANGRGIYVNGSYIITARDYMPR 993
 QY 49 RETI-EEELTGTRFDLQLOVNLITLSGHIDEDD---DEFTLSWS 88
 DB 994 AITAGDVVILTSQANYSVNKTIVTFVNDNDDFENDELSKWN 1038

RESULT 10
 ID W34540
 AC W34540 standard; protein; 394 AA.
 DT 26-MAR-1998 (first entry)
 DE S-adenosylmethionine synthase 1.
 KW S-adenosylmethionine synthase 1; sam1; barley; alkali resistant plant.
 OS Hordeum vulgare.
 PN J09313186-A.
 PD 09-DEC-1997.
 PF 28-MAY-1996; 133406.
 PR 28-MAY-1996; JP-133406.
 PA (NIOC) NIPPON OIL CO LTD.
 DR WPI: 98-080077/08.
 DR N-PSDB; T99141.
 PT S-adenosyl-methionine synthase gene - useful in producing plants
 PT resistant to alkaline soil
 PS Claim 1; Page 5-6; 13pp; Japanese.
 CC This sequence represents the S-adenosylmethionine synthase 1 (sam1)
 CC protein. The DNA encoding this sequence may be used in producing plants
 CC which are resistant to alkaline soil.
 SQ Sequence 394 AA;

Query Match 14.1%; Score 63.5; DB 1; Length 394;
 Best Local Similarity 30.2%; Pred. No. 3.4;
 Matches 32; Conservative 9; Mismatches 34; Indels 31; Gaps 6;

QY 2 SNVFIAFQANED-----SRPIVD---ALVADNPRVAVVSPGMYKIDAPDLRTIRRET 51
 DB 283 SGAYIARQAASIIASGLARRCIVQISYAIYVPEPLSVFVDSYGTGKI--PDR-----E 334
 QY 52 IEELTGTRFDLQ--QLQVNL-----IFLSGHIDEDDDEFT 84
 DB 335 ILKLVKNEFDFRPGMITINLDLKKGNRFRKTAAYGHFGRDDADFT 380

RESULT 11
 ID W34541
 AC W34541 standard; protein; 394 AA.
 DT 26-MAR-1998 (first entry)
 DE S-adenosylmethionine synthase 2.
 KW S-adenosylmethionine synthase 2; sam2; barley; alkali resistant plant.
 OS Hordeum vulgare.
 PN J09313186-A.
 PD 09-DEC-1997.
 PF 28-MAY-1996; 133406.
 PR 28-MAY-1996; JP-133406.
 PA (NIOC) NIPPON OIL CO LTD.
 DR WPI: 98-080077/08.
 DR N-PSDB; T99142.
 PT S-adenosyl-methionine synthase gene - useful in producing plants
 PT resistant to alkaline soil

```
PS Claim 2; Page 6-7; 13pp; Japanese.
CC This sequence represents the S-adenosylmethionine synthase 2 (sam2)
CC protein. The DNA encoding this sequence may be used in producing plants
CC which are resistant to alkaline soil.
SQ Sequence 394 AA;

Query Match 14.1%; Score 63.5; DB 1; Length 394;
Best Local Similarity 30.2%; Pred. No. 3.4;
Matches 32; Conservative 9; Mismatches 34; Indels 31; Gaps 6;

QY 2 SNVFIARQANED-----SRPIVD---AIVADNPRAVWVSPGMVKIDAPDRLTIRRET 51
Db 283 SGATIAQQAASIIASGLARLCIVQISYAIGVPEPLSVFVDSYGTGKI--PDR-----E 334
QY 52 IEELTGTRFDLQ--QLOVNL-----ITLSGHIDEDEDDFT 84
Db 335 ILKLVKENFDRPGMIINLDLKKGNRFKTAAYGHGRDDAFT 380

RESULT 12
W34542
ID W34542 standard; protein; 394 AA.
AC W34542;
DE 26-MAR-1998 (first entry)
DE S-adenosylmethionine synthase 3.
KW S-adenosylmethionine synthase 3; sam3; barley; alkali resistant plant.
OS Hordeum vulgare.
PN J09313186-A.
PD 09-DEC-1997.
PF 28-MAY-1996; 133406.
PR 28-MAY-1996; JP-133406.
PA (NIOC ) NIPPON OIL CO LTD.
DR WPI; 98-080077/08.
DR N-PSDB; T99143.
PT S-adenosyl-methionine synthase gene - useful in producing plants
PT resistant to alkaline soil
PS Claim 3; Page 7-8; 13pp; Japanese.
CC This sequence represents the S-adenosylmethionine synthase 3 (sam3)
CC protein. The DNA encoding this sequence may be used in producing plants
CC which are resistant to alkaline soil.
SQ Sequence 394 AA;

Query Match 14.1%; Score 63.5; DB 1; Length 394;
Best Local Similarity 30.2%; Pred. No. 3.4;
Matches 32; Conservative 9; Mismatches 34; Indels 31; Gaps 6;

QY 2 SNVFIARQANED-----SRPIVD---AIVADNPRAVWVSPGMVKIDAPDRLTIRRET 51
Db 283 SGATIAQQAASIIASGLARLCIVQISYAIGVPEPLSVFVDSYGTGKI--PDR-----E 334
QY 52 IEELTGTRFDLQ--QLOVNL-----ITLSGHIDEDEDDFT 84
Db 335 ILKLVKENFDRPGMIINLDLKKGNRFKTAAYGHGRDDAFT 380

RESULT 13
R11349
ID R11349 standard; Protein; 406 AA.
AC R11349;
DE 05-JUN-1991 (first entry)
DE Cytochrome enzyme P450SU1.
KW Cytochrome P450; P450SU1; P450SU2; herbicide resistance.
OS Streptomyces griseolus.
PN W09103561-A.
PD 21-MAR-1991.
PF 27-AUG-1990; U04785.
PR 11-SEP-1989; US-405605.
PR 12-JAN-1990; US-464499.
PR 23-AUG-1990; US-569781.
PA (DUPO ) DU FONT DE NEMOURS CO.
PI Dean C, Harder PA, Leto KU, Lichtner FT, Odell JT;
```

```
PI O'Keefe DP, Omer CA, Romesser JA;
DR WPI; 91-102077/14.
DR N-PSDB; Q11126.
PT DNA encoding cytochrome P450 enzymes - and electron donating
PT iron-sulphur proteins, used to confer herbicide resistance to
PT plants and microorganisms
PS Claim 13; page 151; 224pp; English.
CC This cytochrome P450 enzyme, P450SU1 is expressed alongside the iron
CC sulphur protein P450-B, by a DNA sequence contained in a recombinant
CC plasmid. Host Streptomyces species are transformed with the plasmid
CC and are used to coat a plant seed to transform the plant. The res-
CC ultant transformants are resistant to herbicides.
CC See also Q11127.
SQ Sequence 406 AA;

Query Match 13.9%; Score 62.5; DB 1; Length 406;
Best Local Similarity 35.5%; Pred. No. 4.8;
Matches 22; Conservative 9; Mismatches 28; Indels 3; Gaps 3;

QY 19 DAIVADNPR-AVWVESP-GMVKIDAPDRLTIRRETIEELTGTRFDLQLOVNLITLSGHI 76
Db 76 DNFPATSPRFAVRESPOAFGLDPPEHCTRRRTISETVKRIKGMPEVEV-VHGEL 134
QY 77 DE 78
Db 135 DE 136

RESULT 14
R61118
ID R61118 standard; Protein; 322 AA.
AC R61118;
DE 07-APR-1995 (first entry)
DE Yeast presequence COX IV-wheat COX II (unedited) fusion protein.
KW Yeast cytochrome oxidase subunit IV presequence; cox IV; cox II;
KW wheat cytochrome oxidase subunit II; chimeric gene;
KW unedited mitochondrial gene; transgenic plant; male-sterile plant;
KW transgene; fusion protein.
OS Chimeric Saccharomyces cerevisiae.
OS Chimeric Triticum sp.
FH Key Location/Qualifiers
FT peptide 1..62
FT /note= "yeast COX IV amino acids 1-62"
FT protein 63..322
FT /label= COX II
FT /note= "unedited version"
PN W09418334-A.
PD 18-AUG-1994.
PF 15-FEB-1994; F00162.
PR 15-FEB-1993; FR-001650.
PA (CNRS ) CNRS CENT NAT RECH SCI.
PI Araya A, Mouras A;
DR WPI; 94-279755/34.
DR N-PSDB; Q70275.
PT Transgenic plant contg non-edited mitochondrial gene - linked to
PT transfer sequence, able to impart male sterility without altering
PT other characteristics, also antisense constructs for restoring
PT fertility
PS Claim 4; page 32-33; 64pp; French.
CC The COX II from wheat is fused to codons 1-62 of the subunit IV of
CC yeast cytochrome oxidase. The COX IV transfer sequence directs the
CC COX II to mitochondria. "Editing" the COX II gene results in
CC changes to 16 amino acids, compared to the unedited version. The
CC unedited COX II is preferred because it results in male-sterility
CC by inhibiting pollen production in transgenic plants.
SQ Sequence 322 AA;

Query Match 13.7%; Score 62; DB 1; Length 322;
Best Local Similarity 32.1%; Pred. No. 4.1;
Matches 27; Conservative 12; Mismatches 25; Indels 20; Gaps 6;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:30:02 ; Search time 58.11 seconds
(without alignments)
47.462 Million cell updates/sec

Title: US-09-430-029-4
Perfect score: 451

Sequence: 1 MSNVFIAFOANEDSRPIVDA.....ITISGHIDEDEFTLSWSH 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	54.8	90	1 DMPM_PSESP	P19731 pseudomonas
2	122	27.1	102	1 TMOD_PSEME	Q00459 pseudomonas
3	69	15.3	822	1 MUS2_SYNY3	P73625 synechocyst
4	66.5	14.7	2481	1 UN52_CAEEL	Q06561 caenorhabdi
5	66	14.6	138	1 MMOB_METTR	P27356 methylosinu
6	66	14.6	181	1 CHMU_ERWHE	P42517 erwinia her
7	66	14.6	619	1 ELFL_HUMAN	P32519 homo sapien
8	65.5	14.5	392	1 METK_MESCR	P93254 mesembryant
9	65	14.4	1162	1 VGL2_IBVM	P12651 avian infec
10	65	14.4	1232	1 Y005_CAEEL	P34643 caenorhabdi
11	64.5	14.3	316	1 CORA_ECOLI	P27841 escherichia
12	64.5	14.3	452	1 KICO_MOUSE	Q61414 mus musculu
13	64	14.2	1162	1 VGL2_IBVB	P11223 avian infec
14	63.5	14.1	224	1 YDHC_BACSU	O03494 bacillus su
15	63.5	14.1	394	1 METK_HORVU	P50299 hordeum vul
16	63.5	14.1	420	1 DODA_ECOLI	P00861 escherichia
17	63	14.0	366	1 METK_PEA	P43612 pisum sativ
18	63	14.0	828	1 YEA4_YEAST	P43585 saccharomyc
19	62.5	13.9	205	1 CMEL_BACSU	P39694 bacillus su
20	62.5	13.9	405	1 CXPE_STRGO	P18326 streptomyce
21	62.5	13.9	964	1 Y0Y1_CAEEL	Q09560 caenorhabdi
22	62	13.7	85	1 Y00G_BPT4	P33285 bacterioph
23	62	13.7	145	1 METL_PETCR	P31156 petrosellinu
24	62	13.7	395	1 METK_POPDE	P47916 populus del
25	62	13.7	612	1 ELFL_MOUSE	Q60775 mus musculu
26	62	13.7	897	1 CRFB_HUMAN	P32927 homo sapien
27	61.5	13.6	254	1 PMFD_PROMI	P53520 proteus mir
28	61.5	13.6	411	1 Y223_MYCGE	P47465 mycoplasma
29	61	13.5	393	1 METL_LYCES	P43281 lycopersico
30	61	13.5	444	1 N1FN_ANASP	Q44145 anabaena sp
31	61	13.5	543	1 CH60_BACSU	P28598 bacillus su
32	61	13.5	583	1 Y40C_RHISN	P55624 rhizobium s
33	60.5	13.4	445	1 PAFL_YEAST	P38351 saccharomyc

34	60.5	13.4	1050	1 ULK1_HUMAN	O75385 homo sapien
35	60.5	13.4	1154	1 VGL2_IBVD2	P12722 avian infec
36	60.5	13.4	1163	1 VGL2_IBV6	P05135 avian infec
37	59.5	13.2	597	1 TREE_EMENI	O42777 emericella
38	59.5	13.2	748	1 METK_PPTCR	P31155 petrosellinu
39	59	13.1	234	1 METK_PPTCR	P31155 petrosellinu
40	59	13.1	376	1 KAPR_CAEEL	P30625 caenorhabdi
41	59	13.1	385	1 SUCC_AQUAE	O67546 aquifex aeo
42	59	13.1	421	1 FTSZ_HAETIN	P45069 haemophilus
43	59	13.1	1423	1 FRUA_STRMU	Q03174 streptococc
44	58.5	13.0	298	1 CBBR_RHORU	P52595 rhodospiril
45	58.5	13.0	393	1 METK_MUSAC	O22338 musa acumin

ALIGNMENTS

RESULT 1
DMPM_PSESP
ID DMPM_PSESP STANDARD; PRT; 90 AA.
AC P19731:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PHENOL HYDROXYLASE P2 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE
DE P2 COMPONENT).
GN DMPM OR PHEA3.
OS Pseudomonas sp. (strain CF600).
OG Plasmid pvii150.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91072230.
RA Nordlund I., Powlowski J., Shingler V.;
RT "complete nucleotide sequence and polypeptide analysis of
RT multicomponent phenol hydroxylase from Pseudomonas sp. strain
RT CF600.";
RL J. Bacteriol. 172:6826-6833(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BH;
RA Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,
RA Fujita M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE: 97164903.
RA Qian H., Edlund U., Powlowski J., Shingler V., Sethson I.;
RT "Solution structure of phenol hydroxylase protein component P2
RT determined by NMR spectroscopy.";
RL Biochemistry 36:495-504(1997).
CC !- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES. P2 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR
CC IN VITRO PHENOL HYDROXYLASE ACTIVITY.
CC !- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) + H(2)O.
CC !- COFACTOR: FAD, AND REQUIRES FE(+2) FOR ACTIVITY.
CC !- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC !- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED
CC BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60276; AAA25941.1; -;
DR EMBL; D28864; BAA06016.1; -;
DR PIR; C37831; C37831.
DR PDB; 1HQI; 23-DEC-96.

RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
 RT "Products of the unc-52 gene in *Caenorhabditis elegans* are homologous
 RT to the core protein of the mammalian basement membrane heparan
 RL sulfate proteoglycan.";
 CC Genes Dev. 7:1471-1484(1993).
 CC -1- FUNCTION: PROBABLE ROLE IN MYOFILAMENT ASSEMBLY AND/OR ATTACHMENT
 CC OF THE MYOFILAMENT LATTICE TO THE CELL MEMBRANE. UNC-52 MAY BE AN
 CC EXTRACELLULAR ANCHOR FOR INTEGRIN RECEPTORS IN MUSCLE.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- ALTERNATIVE PRODUCTS: UNC-52 PRODUCES AT LEAST THREE POLYPEPTIDES:
 CC ONE VERY SHORT FORM AND TWO LONG FORMS.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANE OF ALL
 CC CONTRACTILE TISSUES. IT IS CONCENTRATED OVER MUSCLE DENSE BODIES
 CC AND M-LINES WHICH ARE ASSOCIATED WITH BETA-INTEGRIN.
 CC -1- DEVELOPMENTAL STAGE: SYNTHESIZED EARLY IN EMBRYOGENESIS.
 CC -1- SIMILARITY: CONTAINS 3 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 16 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: TO PERLECAN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L13458; AAA28156.1; --
 DR HSP; P01130; IAJJ.
 DR PFAM; PF00047; ig; 16.
 DR PFAM; PF00052; laminin_B; 2.
 DR PFAM; PF00053; laminin_EGF; 5.
 DR PFAM; PF00057; ldl_recept.a; 3.
 DR PROSITE; PS00022; EGF_1; 4.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01209; LDLRA_1; 3.
 DR PROSITE; PS01209; LDLRA_2; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
 KW Basement membrane; Extracellular matrix; Alternative splicing;
 KW Laminin EGF-like domain.
 FT SIGNAL 1 18
 FT CHAIN 19 2481
 FT DOMAIN 19 148
 FT DOMAIN 149 188
 FT DOMAIN 190 229
 FT DOMAIN 233 272
 FT DOMAIN 273 363
 FT DOMAIN 384 431
 FT DOMAIN 432 441
 FT DOMAIN 442 632
 FT DOMAIN 633 665
 FT DOMAIN 673 719
 FT DOMAIN 720 729
 FT DOMAIN 730 920
 FT DOMAIN 921 953
 FT DOMAIN 954 1003
 FT DOMAIN 1010 1059
 FT DOMAIN 1060 1110
 FT DOMAIN 1133 1216
 FT DOMAIN 1227 1311
 FT DOMAIN 1318 1403
 FT DOMAIN 1415 1499
 FT DOMAIN 1507 1592
 FT DOMAIN 1598 1682
 FT DOMAIN 1699 1785
 FT DOMAIN 1794 1878
 FT DOMAIN 1887 1976
 FT DOMAIN 1977 2070
 FT DOMAIN 2078 2162
 FT DOMAIN 2174 2260

FT DOMAIN 2267 2347
 FT DOMAIN 2354 2436
 FT DISULFID 66 114
 FT DISULFID 149 161
 FT DISULFID 156 174
 FT DISULFID 168 183
 FT DISULFID 190 202
 FT DISULFID 197 215
 FT DISULFID 209 224
 FT DISULFID 233 246
 FT DISULFID 240 259
 FT DISULFID 253 268
 FT DISULFID 954 963
 FT DISULFID 956 970
 FT DISULFID 973 982
 FT DISULFID 985 1001
 FT DISULFID 1010 1020
 FT DISULFID 1012 1026
 FT DISULFID 1029 1038
 FT DISULFID 1041 1057
 FT DISULFID 1060 1068
 FT DISULFID 1062 1078
 FT DISULFID 1081 1090
 FT DISULFID 1093 1108
 FT DISULFID 1151 1199
 FT DISULFID 1337 1383
 FT DISULFID 1434 1480
 FT DISULFID 1526 1572
 FT DISULFID 1617 1662
 FT DISULFID 1718 1766
 FT DISULFID 1813 1860
 FT DISULFID 1906 1953
 FT DISULFID 1997 2052
 FT DISULFID 2098 2146
 FT DISULFID 2194 2241
 FT DISULFID 2283 2328
 FT DISULFID 2373 2419
 FT CARBOHYD 1421 1421
 SQ SEQUENCE 2481 AA; 270812 MW; E47FBE2B310F5031 CRC64;
 Query Match 14.7%; Score 66.5; DB 1; Length 2481;
 Best Local Similarity 28.4%; Pred. NO. 60;
 Matches 21; Conservative 14; Mismatches 24; Indels 15; Gaps 3;
 QY 16 PIVDAIVADNPRAVYVESPGMKIDAPDELITRRTTIBELTGTREDLQ-----QLQVN 68
 Db 1577 PTTNRPVSNPRAVIVKSPIRPIIDPAE-----QTVPE--GSPFKIRCYVPGHPSVOLT 1628
 QY 69 LITLSGHDEDDDE 82
 Db 1629 FRVYSGQLNEDADE 1642
 RESULT 5
 MMOV METTR STANDARD; PRT; 138 AA.
 AC P27356;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE 01-AUG-1992 (Rel. 23, Last annotation update)
 DE METHANE MONOOXYGENASE REGULATORY PROTEIN B.
 GN MMOV.
 OS Methylosinus trichosporium.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Methylocystaceae; Methylosinus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 91251762.
 RA Cardy D.L.N., Laidler V., Salmond G.P.C., Murrell J.C.;
 RT "Molecular analysis of the methane monooxygenase (MMO) gene cluster
 RT of *Methylosinus trichosporium* OB3b.";
 RL Mol. Microbiol. 5:335-342(1991).


```
Db 18 EESQPLNAVWID-----LVEPDDDERLRVQSELGQSL-ATRPELEDIEA----- 61
QY 72 LSGHDEDD 81
Db 62 -SARFFEDD 70

RESULT 12
K1CO_MOUSE STANDARD; PRT; 452 AA.
AC Q61414;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE KERATIN, TYPE I CYTOSKELETAL 15 (CYTOKERATIN 15) (K15) (CK 15).
GN KRT15 OR KRT1-15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=LIVER;
RX MEDLINE; 94171037.
RA Nozaki M., Mori M., Matsushiro A.;
RT "The complete sequence of the gene encoding mouse cytokeratin 15.";
RL Gene 138:197-200(1994).
CC -1- FUNCTION: THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFIBRILLAR
CC KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II (NEUTRAL TO
CC BASIC; 56-70 KDA) [K1 TO K8]. BOTH A BASIC AND AN ACIDIC KERATIN
CC ARE REQUIRED FOR FILAMENT ASSEMBLY.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D16313; BAA03821.1; -.
CC MGD; MGI:96689; KRT1-15.
CC PFAM; PF00038; filament; 1.
CC PROSITE; PS00226; IF; 1.
CC
CC Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin.
FT DOMAIN 1 97 HEAD.
FT DOMAIN 98 406 ROD.
FT DOMAIN 407 452 TAIL.
FT DOMAIN 98 133 COIL 1A.
FT DOMAIN 134 152 LINKER 1.
FT DOMAIN 153 244 COIL 1B.
FT DOMAIN 245 264 LINKER 12.
FT DOMAIN 265 406 COIL 2.
FT DOMAIN 36 39 POLY-GLY.
FT DOMAIN 421 424 POLY-GLY.
FT DOMAIN 425 430 POLY-SER.
SQ SEQUENCE 452 AA; 49159 MW; C431310A7BB408FD CRC64;

Query Match 14.3%; Score 64.5; DB 1; Length 452;
Best Local Similarity 26.0%; Pred. NO.13;
Matches 25; Conservative 19; Mismatches 33; Indels 19; Gaps 4;

QY 8 FOANEDSRPIVDAIVADNPRAVV-VESPGMV-----KIDAPDLITR-----RETI 52
DB 150 PTMEEIRDKILAAATIDRSRVVLEIDNARLAADDFRLKYENELTLRQGVLEADINGLRRLV 209
QY 53 DELTGTREFLD---OOLQVNLITLSLGHDEDDDEFT 84
DB 210 DELTARTDLEMQLEQLNEELAYLAKKNHEEMKEFS 245
```

```
RESULT 13
VGL2_IBVB STANDARD; PRT; 1162 AA.
ID VGL2_IBVB STANDARD; PRT; 1162 AA.
AC P11223; P05134;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN S.
OS Avian infectious bronchitis virus (strain Beaudette) (IBV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85159540.
RA Binns M.M., Boursnell M.E.G., Cavanagh D., Pappind D.J.C.,
RA Brown T.D.K.;
RT "Cloning and sequencing of the gene encoding the spike protein of the
RT coronavirus IBV.";
RL J. Gen. Virol. 66:719-726(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87085499.
RA Binns M.M., Boursnell M.E.G., Tomley F.M., Brown D.K.;
RT "Comparison of the spike precursor sequences of coronavirus IBV
RT strains M41 and 6/82 with that of IBV Beaudette.";
RL J. Gen. Virol. 67:2825-2831(1986).
CC -1- FUNCTION: THE PELOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M95169; AAA70235.1; -.
CC EMBL; X02342; CAA26201.1; -.
CC DR PFAM; PF01600; Corona_S1; 1.
CC DR PFAM; PF01601; Corona_S2; 1.
CC
CC Glycoprotein; Envelope protein; Transmembrane; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1162 E2 GLYCOPROTEIN.
FT CHAIN 19 537 SPIKE PROTEIN S1.
FT CHAIN 538 1162 SPIKE PROTEIN S2.
FT DOMAIN 1120 1137 CYS-RICH.
FT CARBOHYD 51 51 POTENTIAL.
FT CARBOHYD 77 77 POTENTIAL.
FT CARBOHYD 103 103 POTENTIAL.
FT CARBOHYD 144 144 POTENTIAL.
FT CARBOHYD 163 163 POTENTIAL.
FT CARBOHYD 178 178 POTENTIAL.
FT CARBOHYD 212 212 POTENTIAL.
FT CARBOHYD 237 237 POTENTIAL.
FT CARBOHYD 247 247 POTENTIAL.
FT CARBOHYD 264 264 POTENTIAL.
FT CARBOHYD 276 276 POTENTIAL.
FT CARBOHYD 306 306 POTENTIAL.
FT CARBOHYD 425 425 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 513 513 POTENTIAL.
FT CARBOHYD 530 530 POTENTIAL.
FT CARBOHYD 579 579 POTENTIAL.
FT CARBOHYD 591 591 POTENTIAL.
FT CARBOHYD 669 669 POTENTIAL.
FT CARBOHYD 676 676 POTENTIAL.
FT CARBOHYD 714 714 POTENTIAL.
FT CARBOHYD 947 947 POTENTIAL.
FT CARBOHYD 960 960 POTENTIAL.
FT CARBOHYD 979 979 POTENTIAL.
```

```

FT CARBOHYD 1014 1014 POTENTIAL.
FT CARBOHYD 1038 1038 POTENTIAL.
FT CARBOHYD 1051 1051 POTENTIAL.
FT CARBOHYD 1074 1074 POTENTIAL.
SQ SEQUENCE 1162 AA; 128046 MW; 0BAAD58113C9EBD5 CRC64;

Query Match 14.1%; Score 64; DB 1; Length 1162;
Best Local Similarity 26.7%; Pred. No. 45;
Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

QY 4 VFAFAQANEDSRPIVDAIVA-----DNPRVVVSPGMVKIDAPDLRTIR 48
Db 934 VFHFSTPDSFVNVAIVGFCVKPANASQYAIVPANGRGIFQVNGSYITARDMTYMR 993
QY 49 RETI-EEITGRFDLOQLVNLTLSGHIDEDD---DEFTLSWS 88
Db 994 AITAGDVTTTSCQANVSVNKIVITTFVNDDEDFNDELKWN 1038

RESULT 14
YDHC_BACSU STANDARD; PRT; 224 AA.
ID YDHC_BACSU STANDARD; PRT; 224 AA.
AC C05494;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHEICAL TRANSCRIPTIONAL REGULATOR IN DINE-PROB INTERGENIC REGION.
GN YDHC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE; 97346038.
RA Sadaie Y., Yata K., Fujita M., Itaya M., Kasahara Y.,
RA Ogasawara N.;
RT "Nucleotide sequence and analysis of the phoB-rne-groESL region of
RT the Bacillus subtilis chromosome.";
RL Microbiology 143:1861-1866(1997).
CC -1- SIMILARITY: BELONGS TO THE GNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D88802; BAA19694.1; -
CC DR EMBL; Z99107; CAB12389.1; -
CC DR SUBTILIST; BG12180; YDHC.
CC DR PFAM; PF00392; gnr; 1.
CC DR PRINTS; PR00035; HTHGNTR.
CC DR PROSITE; PS00043; HTH_GNR_FAMILY; 1.
CC KW Hypothetical protein; Transcription regulation; DNA-binding.
CC FT DNA_BIND 37 56 H-T-H MOTIF (POTENTIAL).
CC SQ SEQUENCE 224 AA; 26396 MW; 5C41A9E1553B592D CRC64;

Query Match 14.1%; Score 63.5; DB 1; Length 224;
Best Local Similarity 31.7%; Pred. No. 7;
Matches 19; Conservative 14; Mismatches 18; Indels 9; Gaps 3;

QY 30 VVSPGMVKIDAPDLRTIRRET---HELTGTRFDLOQLVNLTLSGHIDE-DDDEFTL 85
Db 57 LLEKDGLLKADRRNGFSITLAKDVDEIYKIRPLEQAVELY-----IDEADEELTI 111

RESULT 15

```

```

METK_HORVU STANDARD; PRT; 394 AA.
ID METK_HORVU STANDARD; PRT; 394 AA.
AC P50299;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE S-ADENOSYLMETHIONINE SYNTHETASE 1 (EC 2.5.1.6) (METHIONINE
DE ADENOSYLTRANSFERASE 1) (ADOMET SYNTHETASE 1).
GN SAM1.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RA Mori S., Takizawa R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 285-322 FROM N.A.
RC TISSUE=ROOT;
RA Mori S., Takizawa R., Nakanishi H.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC METHIONINE AND ATP.
CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + H(2)O = ORTHOPHOSPHATE +
CC PYROPHOSPHATE + S-ADENOSYL-L-METHIONINE.
CC -1- PATHWAY: ACTIVATED METHYL CYCLE.
CC -1- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D63835; BAA09895.1; -
CC DR EMBL; D49655; BAA08531.1; -
CC DR HSP; P04384; IXRC.
CC DR PFAM; PF00438; S-Adomet_synt; 1.
CC DR PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
CC DR PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
CC KW Transferase; One-carbon metabolism; Multigene family; ATP-binding.
CC FT NP_BIND 121 126 ATP (POTENTIAL).
CC FT BINDING 149 149 ATP (POTENTIAL).
CC SQ SEQUENCE 394 AA; 42841 MW; 2ED786F333530013 CRC64;

Query Match 14.1%; Score 63.5; DB 1; Length 394;
Best Local Similarity 30.2%; Pred. No. 14;
Matches 32; Conservative 9; Mismatches 34; Indels 31; Gaps 6;

QY 2 SNVFIATQANED-----SRPIVD---AIVADNPRAVVPSPGMVKIDAPDLRTIRRET 51
Db 283 SGAYIARQAASIIASGLARRCIQVSYAIGVPEPLSVFVDSYGTGKI--PDR-----E 334
QY 52 IEELTGTRFDLO--OLOVNL-----ITLSGHIDEDEFT 84
Db 335 ILKLIVENDFRPGMITINLDLKKGNRFITAAAYGHRDDADFT 380

```

Search completed: September 26, 2000, 20:30:04
Job time: 428 sec

2

Wed Sep 27 17:17:02 2000

us-09-430-029-4.rsp

Page 9

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:28:52 ; Search time 92.32 seconds
(without alignments)
66.841 Million cell updates/sec

Title: US-09-430-029-4
Perfect score: 451
Sequence: 1 MSNVFTAFQANESRPIVDA.....IILSGHIDEDDEFTLSWSH 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_12:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	354	78.5	89	2 030591	O30591 burkholderi
2	346	76.7	89	2 Q52571	Q52571 pseudomonas
3	326	72.3	97	2 Q92NP5	Q92np5 comamonas t
4	257	57.0	90	2 O84960	O84960 ralstonia s
5	247	54.8	90	2 Q52172	Q52172 pseudomonas
6	243	53.9	91	2 Q52163	Q52163 pseudomonas
7	233	51.7	89	2 Q32430	Q32430 acinetobact
8	232	51.4	89	2 Q43980	Q43980 acinetobact
9	114	25.3	110	2 O87801	O87801 pseudomonas
10	111	24.6	147	2 P95411	P95411 pseudomonas
11	110	24.4	104	2 Q51942	Q51942 burkholderi
12	105	23.3	105	2 Q69181	Q69181 alcaligenes
13	90	20.0	101	2 Q92ET4	Q92et4 xanthobacte
14	84	18.6	104	2 O07071	O07071 burkholderi
15	83.5	18.5	862	2 Q9X8M6	Q9x8m6 streptomyce
16	77	17.1	1684	4 Q75218	Q75218 homo sapien
17	77	17.1	3830	4 Q9Y6H4	Q9y6h4 homo sapien
18	77	17.1	3859	4 Q9F631	Q9y631 homo sapien
19	70.5	15.6	174	1 O28973	O28973 archaeoglob

20	67.5	15.0	410	2	Q59831	streptomyce
21	66.5	14.7	2295	5	Q9XTD2	caenorhabdi
22	66.5	14.7	2482	5	O18263	caenorhabdi
23	66.5	14.7	3375	5	Q9XTI5	caenorhabdi
24	66	14.6	393	1	O30028	caenorhabdi
25	66	14.6	649	2	P73736	archaeoglob
26	65	14.4	263	5	O76870	synecocyst
27	65	14.4	281	2	Q9ZHS9	drosofila
28	65	14.4	1153	12	Q66178	streptomyce
29	65	14.4	1159	12	Q82624	avian infec
30	64	14.2	352	3	Q9Y887	suillus bov
31	64	14.2	395	4	Q9Y308	homo sapien
32	63.5	14.1	983	2	O30988	pseudomonas
33	63	14.0	138	2	O06118	methyloyst
34	63	14.0	557	2	Q00876	renibacteri
35	63	14.0	690	2	Q9X6R4	aeromonas p
36	62.5	13.9	139	1	O27443	methanobact
37	62.5	13.9	1541	2	O54298	streptomyce
38	62.5	13.9	4077	2	O52820	amycolatops
39	62	13.7	166	7	Q95517	macaca mula
40	62	13.7	274	8	Q37697	zea mays (m
41	62	13.7	354	7	Q95513	macaca fasc
42	62	13.7	354	7	Q95514	macaca fasc
43	62	13.7	811	12	O36185	plautia sta
44	62	13.7	941	11	O63128	rattus norv
45	62	13.7	1653	2	P76578	escherichia

ALIGNMENTS

RESULT 1
Q30591 PRELIMINARY; PRT; 89 AA.
ID O30591;
AC O30591;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CRPC PROTEIN.
GN CRPC.
OS Burkholderia pickettii (Pseudomonas pickettii).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PK01;
RA OLSEN R.H., KUKOR J.J., BYRNE A.M., JOHNSON G.R.;
RL J. Ind. Microbiol. 0:0-0(1997).
DR EMBL; AF012632; AAB67107.1; -.
DR HSSP; P19731; 1HQI.
SQ SEQUENCE 89 AA; 10055 MW; BD74CF41 CRC32;

Query Match 78.5%; Score 354; DB 2; Length 89;
Best Local Similarity 76.4%; Pred. No. 1.4e-29;
Matches 68; Conservative 7; Mismatches 14; Indels 0; Gaps 0;
QY 1 MSNVFTAFQANESRPIVDAIVADNPRAVVEPSGMVKIDAPRLIRRETIETGTGRF 60
DB 1 MSQVFTAFQANESRPIVDAIVADNPRAVVEPSGMVKIDAPRLIRRETIETGTGRF 60
QY 61 DLQQLQVNLITLSGHIDEDDEFTLSWSH 89
DB 61 DLQQLHVNLTLSGHIDEDDEFTLSWSH 89

RESULT 2
Q52571 PRELIMINARY; PRT; 89 AA.
ID Q52571;
AC Q52571;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

```

DE  TOLUENE/BENZENE-2-MONOOXYGENASE (TBMA-TMMF) GENES, COMPLETE CDS'S
DE  (TBMA-TMMF).
GN  TBMC.
OS  Pseudomonas sp.
OC  Bacteria; Proteobacteria.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-JS150;
RX  MEDLINE; 96035667.
RA  JOHNSON G.R., OLSEN R.H.;
RT  "Nucleotide sequence analysis of genes encoding a toluene/benzene-2-
RL  monooxygenase from Pseudomonas sp. strain JS150.";
RL  Appl. Environ. Microbiol. 61:3336-3346(1995).
DR  EMBL; L40033; AAA88458.1; -.
DR  HSSP; P19731; IHQI.
SQ  SEQUENCE 89 AA; 10029 MW; F54D09A2 CRC32;

Query Match 76.7%; Score 346; DB 2; Length 89;
Best Local Similarity 75.3%; Pred. No. 9.3e-29;
Matches 67; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY  1 MSNVFIQFQANDSRPIVDIAVADNPRAVVPSPGMVKIDAPDRITIRRETIETLTGTRF 60
DB  1 MSQVFIAQFQNEESRPVDAILLADNTHAQVYSPGLVKIDAPGRITIRRETIETLTGTRF 60
QY  61 DLQQLQVNLITLSGHIDDDDEFTLSWS 89
DB  61 DLQQLHVNILVLSGHIDDDDDQLTISWQH 89

RESULT 3
Q92NP5 PRELIMINARY; PRT; 97 AA.
AC Q92NP5;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN APHM.
OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TA441;
RX MEDLINE; 99018839.
RA ARAI H., AKAHIRA S., OHISHI T., MAEDA M., KUDO T.;
RT "Adaptation of Comamonas testosteroni TA441 to utilize phenol:
RT organization and regulation of the genes involved in phenol
RT degradation.";
RL Microbiology 144:2895-2903(1998).
DR EMBL; AB006479; BAA34171.1; -.
DR HSSP; P19731; IHQI.
SQ SEQUENCE 97 AA; 10728 MW; 47CD2246 CRC32;

Query Match 72.3%; Score 326; DB 2; Length 97;
Best Local Similarity 69.3%; Pred. No. 1.2e-26;
Matches 61; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY  1 MSNVFIQFQANDSRPIVDIAVADNPRAVVPSPGMVKIDAPDRITIRRETIETLTGTRF 60
DB 9 VSKVFIAQFQNEESRPVDAILLADNTHAQVYSPGLVKIDAPGRITIRRETIETLTGTRF 68
QY  61 DLQQLQVNLITLSGHIDDDDEFTLSWS 88
DB 69 ELQQTQINLVLSGHIDDDDDFTLSWN 96

RESULT 4
O84960 PRELIMINARY; PRT; 90 AA.
ID O84960
AC O84960;

```

```

DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN POXC.
OS Ralstonia sp. E2.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E2;
RX MEDLINE; 98361023.
RA HINO S., WATANABE K., TAKAHASHI N.;
RT "Phenol hydroxylase cloned from Ralstonia eutropha strain E2 exhibits
RT novel kinetic properties.";
RL Microbiology 144:0-0(0).
DR EMBL; AF026065; AAC32454.1; -.
DR HSSP; P19731; IHQI.
SQ SEQUENCE 90 AA; 9956 MW; 8B0F5684 CRC32;

Query Match 57.0%; Score 257; DB 2; Length 90;
Best Local Similarity 57.5%; Pred. No. 1.3e-19;
Matches 50; Conservative 16; Mismatches 21; Indels 0; Gaps 0;

QY  2 SNVFIQFQANDSRPIVDIAVADNPRAVVPSPGMVKIDAPDRITIRRETIETLTGTRF 61
DB 3 ANVYIALQNNDDTRPIIEAIAEANPLAVVSPGMVKIDAPGRITIRRETIETLTGTRF 62
QY  62 LQQLQVNLITLSGHIDDDDEFTLSWS 88
DB 63 LQEIHLNLSLGSNIDETDEAFTLWS 89

RESULT 5
Q52172 PRELIMINARY; PRT; 90 AA.
AC Q52172;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE P2 PROTEIN (EC 1.14.13.7)
DE (PHENOL 2-MONOOXYGENASE P2 COMPONENT).
GN PHHM.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P35X / NCBI 9869;
RX MEDLINE; 95129877.
RA NG L.C., SHINGLER V., SZE C.C., POH C.L.;
RT "Cloning and sequences of the first eight genes of the chromosomally
RT encoded (methyl) phenol degradation pathway from Pseudomonas putida
RT P35X.";
RL Gene 151:29-36(1994).
CC - FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC - CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2)O.
CC - COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC - PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC - SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
DR EMBL; X79063; CAA55662.1; -.
DR HSSP; P19731; IHQI.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron.
SQ SEQUENCE 90 AA; 10519 MW; D9015E61 CRC32;

Query Match 54.8%; Score 247; DB 2; Length 90;

```

```

Best Local Similarity 52.3%; Pred. No. 1.4e-18;
Matches 46; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 2 SNVFIAFQANEDSRPIYDAIVADNPRAVVESPGMKVADPRLTIRRETEELTGTRFD 61
Db 3 SLVYIAFQDNARLYEAIQDNPHAVVQHHPAMIRIEAEKRLIRREIVEENLGRAND 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWSH 89
Db 63 VQEMLVDTVITIGNIDEDDRFVLEWKN 90

RESULT 6
Q52163 PRELIMINARY; PRT; 91 AA.
ID Q52163
AC Q52163;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE P2 PROTEIN (EC 1.14.13.7)
DE {PHENOL 2-MONOOXYGENASE P2 COMPONENT}.
GN PHLC.
OS Pseudomonas putida.
OG Plasmid pPGH1.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE; 95272534.
RA HERMANN H., MUELLER C., SCHWIDT I., MAHNKE J., PETRUSCHKA L.,
RA HAHNKE K.;
RT "Localization and organization of phenol degradation genes of
RT Pseudomonas putida strain H."
RL Mol. Gen. Genet. 247:240-246(1995).
CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2)O.
CC -!- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
DR EMBL; X60765; CAA56742.1; -.
DR HSSP; P19731; 1HQ1.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron; Plasmid.
SQ SEQUENCE 91 AA; 10661 MW; 129BLFDF CRC32;

Query Match 53.9%; Score 243; DB 2; Length 91;
Best Local Similarity 50.0%; Pred. No. 3.7e-18;
Matches 44; Conservative 21; Mismatches 23; Indels 0; Gaps 0;

QY 2 SNVFIAFQANEDSRPIYDAIVADNPRAVVESPGMKVADPRLTIRRETEELTGTRFD 61
Db 3 SLVYIAFQDNARLYEAIQDNPHAVVQHHPAMIRIEAEKRLIRREIVEENLGRAND 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWSH 89
Db 63 VQEMLVDTVITIGNIDEDDRFVLEWKN 90

RESULT 7
Q32430 PRELIMINARY; PRT; 89 AA.
ID Q32430
AC Q32430;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE DMS OXYGENASE COMPONENT.
GN DSOC.

Best Local Similarity 52.3%; Pred. No. 1.4e-18;
Matches 46; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

OS Acinetobacter sp.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Moraxellaceae; Acinetobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20B;
RX MEDLINE; 98005684.
RA HORINOCHI M., KASUGA K., NOJIRI H., YAMANE H., OMORI T.;
RT "Cloning and characterization of genes encoding an enzyme which
RT oxidizes dimethyl sulfide in Acinetobacter sp. strain 20B."
RL FEMS Microbiol. Lett. 155:99-105(1997).
DR EMBL; D85083; BAA23332.1; -.
DR HSSP; P19731; 1HQ1.
SQ SEQUENCE 89 AA; 10135 MW; A504444FF CRC32;

Query Match 51.7%; Score 233; DB 2; Length 89;
Best Local Similarity 46.0%; Pred. No. 3.9e-17;
Matches 40; Conservative 22; Mismatches 25; Indels 0; Gaps 0;

QY 2 SNVFIAFQANEDSRPIYDAIVADNPRAVVESPGMKVADPRLTIRRETEELTGTRFD 61
Db 3 SKVYLALQDNDSRYITIEAIEQDNPEATIQLPAMIRVESTGELVRAETVSEKLGQNW 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWS 88
Db 63 IQELQLNMITLGGNVDEDDDSFTLKWN 89

RESULT 8
Q43980 PRELIMINARY; PRT; 89 AA.
ID Q43980
AC Q43980;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PHENOLHYDROXYLASE COMPONENT.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Moraxellaceae; Acinetobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB8250;
RX MEDLINE; 96154937.
RA EHRT S., SCHIRMER F., HILLEN W.;
RT "Genetic organization, nucleotide sequence and regulation of
RT dioxygenase in Acinetobacter calcoaceticus NCIB8250."
RL Mol. Microbiol. 18:13-20(1995).
DR EMBL; Z36909; CAA85382.1; -.
DR HSSP; P19731; 1HQ1.
SQ SEQUENCE 89 AA; 10136 MW; AA99E933 CRC32;

Query Match 51.4%; Score 232; DB 2; Length 89;
Best Local Similarity 46.0%; Pred. No. 4.9e-17;
Matches 40; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

QY 2 SNVFIAFQANEDSRPIYDAIVADNPRAVVESPGMKVADPRLTIRRETEELTGTRFD 61
Db 3 SKVYLALQDNDSRYITIEAIEQDNPEATIQLPAMIRVESTGELVRAETVSEKLGQNW 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWS 88
Db 63 IQELQLNMITLGGNVDEDDDSFTLKWN 89

RESULT 9
O87801 PRELIMINARY; PRT; 110 AA.
ID O87801
AC O87801;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)

```

```

DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE TOUD.
GN Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OX1;
RX MEDLINE; 98432776.
RA BERTONI G., MARTINO M., GALLI E., BARBIERI P.;
RT "Analysis of the gene cluster encoding toluene/o-xylene monooxygenase
from Pseudomonas stutzeri OX1."
RL Appl. Environ. Microbiol. 64:3626-3632(1998).
DR EMBL; AJ005663; CAA06657.1; -.
DR HSP; P19731; IHQ1.
SQ SEQUENCE 110 AA; 12274 MW; A15B73F3 CRC32;

Query Match 25.3%; Score 114; DB 2; Length 110;
Best Local Similarity 37.3%; Pred. No. 8.6e-05;
Matches 28; Conservative 15; Mismatches 30; Indels 2; Gaps 1;

QY 9 QANEDSRPIVDAIVADNP--RAVVVSPGMVKIDAPDRLTIIRRETIETLTGTRFDLQQLQ 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 26 RAGDVAIVAGEAAEDNPGKEIKVDKLAYVRGAEDILIRKETIEECLGRPFMRQELE 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 VNLITLSGHIDEEDD 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 INLSSFAGIIDMDFD 100

RESULT 10
ID P95411 PRELIMINARY; PRT; 147 AA.
AC P95411;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE BM01 PROTEIN.
GN BM01.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J1104;
RA KITAYAMA A., SUZUKI E., KAWAKAMI Y., NAGAMUNE T.;
RT "Gene organization and low regio specificity in aromatic-ring
RT hydroxylation of a benzene monooxygenase of Pseudomonas aeruginosa
RT J1104."
RL J. Ferment. Bioeng. 82:421-425(1996).
DR EMBL; D83068; BAAL1764.1; -.
SQ SEQUENCE 147 AA; 16261 MW; 84995200 CRC32;

Query Match 24.6%; Score 111; DB 2; Length 147;
Best Local Similarity 33.8%; Pred. No. 0.00025;
Matches 25; Conservative 16; Mismatches 31; Indels 2; Gaps 1;

QY 9 QANEDSRPIVDAIVADNP--RAVVVSPGMVKIDAPDRLTIIRRETIETLTGTRFDLQQLQ 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 RAGELADAVAAFAAREDPNGKEIRVDKRAYLRIDTDDEMIIRRTTIEQALGRPFDMPELE 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 VNLITLSGHIDEEDD 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 INLSSFAGIQTND 137

RESULT 11
Q51942 PRELIMINARY; PRT; 104 AA.
ID Q51942
AC Q51942;

```

```

DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TReMBLrel. 08, Last annotation update)
DE TOJUENE-3 MONOOXYGENASE FERREDOXIN PROTEIN.
GN TBUV.
OS Burkholderia pickettii (Pseudomonas pickettii).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PK01;
RX MEDLINE; 95172404.
RA BYRNE A.M., KUKOR J.J., OLSEN R.H.;
RT "Sequence analysis of the gene cluster encoding toluene-3-
monooxygenase from Pseudomonas pickettii PK01.";
RL Gene 154:65-70(1995).
DR EMBL; U04052; AAB09621.1; -.
KW Monooxygenase.
SQ SEQUENCE 104 AA; 11703 MW; 31CA0AC7 CRC32;

Query Match 24.4%; Score 110; DB 2; Length 104;
Best Local Similarity 32.0%; Pred. No. 0.00021;
Matches 24; Conservative 20; Mismatches 29; Indels 2; Gaps 2;

QY 9 QANEDSRPIVDAIVADNP--RAVVVSP--GMVKIDAPDRLTIIRRETIETLTGTRFDLQQLQ 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 RASITAGVIEAAQEDNPGKSIRIDDKLAYVRIDTDGELLIRATLEALGRPFKASELE 79
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 VNLITLSGHIDEEDD 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 VNLSSFAGRIETD 94

RESULT 12
ID O69181 PRELIMINARY; PRT; 105 AA.
AC O69181;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (TReMBLrel. 07, Last annotation update)
DE PUTATIVE HYDROXYLASE COMPONENT.
GN PHLN.
OS Alcaligenes eutrophus.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JMP134;
RA HARKER A.R., AYOUBI P.J.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065891; AAC77383.1; -.
SQ SEQUENCE 105 AA; 11913 MW; 3971DE3F CRC32;

Query Match 23.3%; Score 105; DB 2; Length 105;
Best Local Similarity 32.0%; Pred. No. 0.00069;
Matches 24; Conservative 17; Mismatches 32; Indels 2; Gaps 1;

QY 9 QANEDSRPIVDAIVADNP--RAVVVSPGMVKIDAPDRLTIIRRETIETLTGTRFDLQQLQ 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 RASITSGVIEAAQEDNPGKEIRVDKRAYVRIDTDGELLIRATLEALGRPFMRSELE 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 VNLITLSGHIDEEDD 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 VNLSSFAGRIETD 95

RESULT 13
Q9ZET4 PRELIMINARY; PRT; 101 AA.
ID Q9ZET4
AC Q9ZET4;
DT 01-MAY-1999 (TReMBLrel. 10, Created)

```

```

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE COUPLING/EFFECTOR PROTEIN.
GN XAMOD.
OS Xanthobacter sp. Py2.
OC Bacteria; Proteobacteria; alpha subdivision; Ancylobacter group;
OC Xanthobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PY2;
RA ZHOU N.Y., CHAN KWO CHION C.K., LEAK D.J.;
RT "The alkene monooxygenase from Xanthobacter Py2 is closely related to
RT aromatic monooxygenase and catalyses aromatic monohydroxylation of
RT benzene, toluene and phenol.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ012090; CAA09914.1; -.
SQ SEQUENCE 101 AA; 11194 MW; F4F59639 CRC32;

Query Match      20.0%; Score 90; DB 2; Length 101;
Best Local Similarity 24.2%; Pred. No. 0.023;
Matches 24; Conservative 19; Mismatches 42; Indels 14; Gaps 3;

QY 1 MSNVFI-----AFQANEDSRIVDAIVADNP--RAVVESPGMKIDAPDRLTTR 48
Db 1 MSNAIVDDMDENLVGVIRAGDLADAVIADNPGKEVHIERGDYVRIHTDRDCRLT 60
QY 49 RETIEELGTRELDLQQLVNLTLSGHIDEDEDEFTLSW 87
Db 61 RASIEQALGRSFVLAIAEAMSFKGRMSSDSE--MRW 97

RESULT 14
ID O07071 PRELIMINARY; PRT; 104 AA.
AC O07071;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE TBHD.
GN TBHD.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AA1;
RA MA Y., HERSON D.S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001356; AAB58743.1; -.
SQ SEQUENCE 104 AA; 11790 MW; 40251740 CRC32;

Query Match      18.6%; Score 84; DB 2; Length 104;
Best Local Similarity 34.3%; Pred. No. 0.098;
Matches 23; Conservative 10; Mismatches 28; Indels 6; Gaps 2;

QY 17 IVDAIV----ADNP--RAVVESPGMKIDAPDRLTIRRETIEELTGTRELDLQQLVNLI 70
Db 24 LVDAFVEARADNPGEICVDKRAYIRIDTGGELILPRETIERALGRFFKMPDLEVELS 83
QY 71 TLSGHID 77
Db 84 SPAGRIE 90

RESULT 15
QYX8M6 PRELIMINARY; PRT; 862 AA.
ID QYX8M6
AC QYX8M6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

```

```

DE PUTATIVE LARGE ATP-BINDING PROTEIN.
GN SCE94.20.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA OLIVER K., HARRIS D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049628; CAB40869.1; -.
KW ATP-binding.
SQ SEQUENCE 862 AA; 94537 MW; 631192DD CRC32;

Query Match      18.5%; Score 83.5; DB 2; Length 862;
Best Local Similarity 28.6%; Pred. No. 1.4;
Matches 18; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 27 RAVVESPGMKIDAPDRLTIRRETIEELTGTRELDLQQLVNLTLSGHIDEDEDEFTLS 86
Db 556 RALLRS-GILREPARGRVDFLHRTFDQYLGARLAVQEMDFOLLVNHHLDEWDDVILLA 614
QY 87 WSH 89
Db 615 LAH 617

Search completed: September 26, 2000, 20:28:54
Job time: 1013 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:30:04 ; Search time 58.11 Seconds
(without alignments)
275.174 Million cell updates/sec

Title: US-09-430-029-5

Perfect score: 2873

Sequence: 1 MDPTLKKLGLKDRYAAMT.....FDGSEDKNFAAWRGQATRN 516

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1846.5	64.3	517	1	DMPN_PSESP
2	374.5	13.0	499	1	TMOA_PSEME
3	234.5	8.9	527	1	MEMA_METCA
4	211	7.3	525	1	TMOA_METTR
5	152	5.3	326	1	TMOE_PSEME
6	115	4.0	852	1	POL_BIVAU
7	107.5	3.7	852	1	POL_BIVJ
8	104	3.6	549	1	AMY_BACST
9	103	3.6	526	1	MLHL_ARATH
10	103	3.6	549	1	MCRA_METTH
11	99	3.4	549	1	MCRA_METTM
12	95	3.3	782	1	OSTA_HAEIN
13	94.5	3.3	1306	1	ACE_HUMAN
14	92.5	3.2	842	1	DMPN_LACDL
15	90.5	3.2	1256	1	MRP_STRSU
16	89.5	3.1	971	1	AGLU_ASPNG
17	89.5	3.1	985	1	ATSB_KLEAE
18	89	3.1	405	1	OSTA_HAEIN
19	89	3.1	670	1	TBUD_BURPL
20	89	3.1	875	1	VP34_YEAST
21	88	3.1	582	1	ASN2_PEA
22	88	3.1	713	1	DC12_ECOLI
23	87.5	3.0	702	1	ARYA_MANSE
24	87.5	3.0	1009	1	AMPN_HELVI
25	87	3.0	781	1	GCSL_CABEL
26	87	3.0	819	1	STL_CHLTR
27	86.5	3.0	666	1	Y032_MYCGE
28	86.5	3.0	815	1	GYRB_MYXXA
29	86	3.0	592	1	INV2_DAUCA
30	85.5	3.0	483	1	PHR_ANAT
31	84.5	2.9	585	1	ASN2_LOTJA
32	84.5	2.9	604	1	IAP1_HUMAN
33	84.5	2.9	723	1	SYM_PYRHO

34	84.5	2.9	1237	1	YDY2_SCHPO
35	84	2.9	535	1	YGIS_ECOLI
36	84	2.9	585	1	ASNL_LOTJA
37	84	2.9	619	1	PPCK_HAECO
38	84	2.9	621	1	HTPG_RICPR
39	84	2.9	757	1	DHET_GLOSU
40	83.5	2.9	605	1	MALZ_ECOLI
41	83	2.9	760	1	SPOT_SYNK3
42	82.5	2.9	557	1	G6PI_KLULA
43	82.5	2.9	585	1	YM67_CAREL
44	82.5	2.9	607	1	SYRC_YEAST
45	82.5	2.9	850	1	CYAA_YERPE

ALIGNMENTS

RESULT 1

DMPN_PSESP

ID DMPN_PSESP STANDARD; PRT; 517 AA.

AC P19732;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE PHENOL HYDROXYLASE P3 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE

DE P3 COMPONENT).

GN DMPN OR PHEM4.

OS Pseudomonas sp. (strain CF600).

OG Plasmid pVil150.

OC Bacteria; Proteobacteria.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 91072230.

RA Nordlund I., Powlowski J., Shingler V.;

RT "Complete nucleotide sequence and polypeptide analysis of

multicomponent phenol hydroxylase from Pseudomonas sp. strain

CF600.";

RL J. Bacteriol. 172:6826-6833(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BH;

RA Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,

Fujita M.;

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED

DERIVATIVES. P3 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR

IN VITRO PHENOL HYDROXYLASE ACTIVITY.

CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) + H(2)O.

CC -!- COFACTOR: FAD, AND REQUIRES FE(+2) FOR ACTIVITY.

CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.

CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED

BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M60276; AAA25942.1; -

DR EMBL; D28864; BAA06017.1; -

DR PIR; D37831; D37831.

KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;

NADP; Flavoprotein; FAD; Iron; Plasmid.

SQ SEQUENCE 517 AA; 60522 MW; 4EED4EB40ED73F9C CRC64;

Query Match 64.3%; Score 1846.5; DB 1; Length 517;
Best Local Similarity 64.8%; Pred. No. 2.8e-145;
Matches 328; Conservative 63; Mismatches 106; Indels 9; Gaps 1;

RA Rosenzweig A.C., Frederick C.A., Lippard S.J., Nordlund P.;
 RT "Crystal structure of a bacterial non-haem iron hydroxylase that
 RL catalyses the biological oxidation of methane.";
 CC Nature 366:537-543(1993).
 CC !- FUNCTION: RESPONSIBLE FOR THE INITIAL OXYGENATION OF METHANE TO
 CC METHANOL IN METHANOTROPHS. IT ALSO CATALYZES THE MONOHYDROXYLATION
 CC OF A VARIETY OF UNACTIVATED ALKENES, ALICYCLIC, AROMATIC AND
 CC HETEROCYCLIC COMPOUNDS.
 CC !- CATALYTIC ACTIVITY: METHANE + NAD(P)H + O(2) = METHANOL +
 CC NAD(P)(+) + H(2)O.
 CC !- COFACTOR: BINDS TWO IRON ATOMS.
 CC !- SUBUNIT: M.CAPSULATUS HAS TWO FORMS OF METHANE MONOOXYGENASE,
 CC A SOLUBLE AND A MEMBRANE-BOUND TYPE. THE SOLUBLE TYPE CONSISTS
 CC OF THREE COMPONENTS (A, B AND C): PROTEIN A, COMPRISING THREE
 CC CHAINS, IN AN ALPHA-2, BETA-2, GAMMA-2 CONFIGURATION, IS A
 CC NONHEME IRON PROTEIN CONTAINING AN UNUSUAL MU-HYDROXO BRIDGE
 CC STRUCTURE AT ITS ACTIVE SITE AND INTERACTS WITH BOTH OXYGEN AND
 CC METHANE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M58499; AAA25385.1; -;
 CC EMBL; M90050; AAB62392.1; -;
 CC PIR; JQ0702; JQ0702.
 CC
 CC Oxidoreductase; Monooxygenase; NADP; One-carbon metabolism; Iron.
 KW ACT_SITE 151 151 POTENTIAL.
 FT METAL 114 114 IRON 1.
 FT METAL 144 144 IRON 1.
 FT METAL 147 147 IRON 1.
 FT METAL 209 209 IRON 2.
 FT METAL 243 243 IRON 2.
 FT METAL 246 246 IRON 2.
 SQ SEQUENCE 527 AA; 60630 MW; 44DBB7A325423049 CRC64;

Query Match 8.9%; Score 254.5; DB 1; Length 527;
 Best Local Similarity 24.2%; Pred. No. 1.3e-13;
 Matches 109; Conservative 65; Mismatches 179; Indels 97; Gaps 22;
 QY 54 DPFLRTMDAYWYQGEKELLYAVID-AFTONNAPFLGVSDARYINALKFLQGVTPLEVL 112
 DB 58 EQFKLIAKEYARMEAVKDERQSGLSQVALTRLNAGVRV-HPKWNTEIMKV 106
 QY 113 AHRGFAHVGRH-----FTG---EGARIACQ-----MQSIDEIRHYQETETHAMSTYKNKFF-- 158
 DB 107 --SNELEVGEYNATAATGMLWDSQAQKNGYLQAQVLDEIRH-----THQCAVYVYFAK 160
 QY 159 NG---FHHSNQWDFR---VWYLSVPKSEFFEDAYSGGFEFLTAVSFSE-----EYVLT 205
 DB 161 NGQDPAGHNDARRTRTIGPLW--KGMKRVFSDFGTSFSG-----DAVECSINLQLVGEACFT 213
 QY 206 NLLFPFMSGAAINGDMSVTVTFGSAQSDSHRMTLGTEICKFLEQDPNNVPIVQRWID 265
 DB 214 NPLIVAVTEWAANGDEITPVPFLSIEFDELHMHANGQTV-VSTANDPASKYINTLDIN 272
 QY 266 KWFNRYGYKLLT-LVAMMDYMWOPKRVMSWRSEWYMA-EONGGALFKDLARYGIREPKGW 323
 DB 273 NAFWTQQKYFTPVGLMFLFYGSKFKVPEWVKTNWVYEDWGGIWLGRGKYGVESPSRL 332
 QY 324 QDACEGKDHISHQAWSTFYGFNAASAFHTWPTDEMGWLSAKYPDSDFDRIYRPRFDHWG 383
 DB 333 KDAQODAYWAHHDLVLLAYALMPTGFFRLALPDQCEMEFWEFANYPGWYD-HYKGIYEW- 390
 QY 384 EQARAGNRFYMKTLPMIQCQOIP-----MLFTEPENGPTKI-----GA- 421
 DB 391 -RARG-----CEDPSSGGIPIIMWFIEENNHPYIDRVSVQVFPFCPSLAKGAS 434

QY 422 --RESNYLGNKFHFCSDHCKDIFDHEPOKY 449
 DB 435 TLRVHEYNGQMHFTSDOWGERWKLAEPEKY 464
 RESULT 4
 MEMA_METTR STANDARD; PRT; 525 AA.
 AC P27353;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE METHANE MONOOXYGENASE COMPONENT A ALPHA CHAIN (EC 1.14.13.25) (METHANE
 DE HYDROXYLASE).
 GN MMOX.
 OS Methylosinus trichosporium.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Methylocystaceae; Methylosinus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OB3B;
 RX MEDLINE; 91251762.
 RA Cardy D.L.N., Laidler V., Salmond G.P.C., Murrell J.C.;
 RT "Molecular analysis of the methane monooxygenase (MMO) gene cluster
 RT of Methylosinus trichosporium OB3B.";
 RL Mol. Microbiol. 5:335-342(1991).
 CC !- FUNCTION: RESPONSIBLE FOR THE INITIAL OXYGENATION OF METHANE TO
 CC METHANOL IN METHANOTROPHS. IT ALSO CATALYZES THE MONOHYDROXYLATION
 CC OF A VARIETY OF UNACTIVATED ALKENES, ALICYCLIC, AROMATIC AND
 CC HETEROCYCLIC COMPOUNDS.
 CC !- CATALYTIC ACTIVITY: METHANE + NAD(P)H + O(2) = METHANOL +
 CC NAD(P)(+) + H(2)O.
 CC !- COFACTOR: BINDS TWO IRON ATOMS.
 CC !- SUBUNIT: M.TRICHOSPORIUM HAS TWO FORMS OF METHANE MONOOXYGENASE,
 CC OF THREE COMPONENTS (A, B AND C): PROTEIN A, COMPRISING THREE
 CC CHAINS, IN AN ALPHA-2, BETA-2, GAMMA-2 CONFIGURATION, IS A
 CC NONHEME IRON PROTEIN CONTAINING AN UNUSUAL MU-HYDROXO BRIDGE
 CC STRUCTURE AT ITS ACTIVE SITE AND INTERACTS WITH BOTH OXYGEN AND
 CC METHANE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X55394; CAA39068.1; -;
 CC PIR; S15207; S15207.
 KW Oxidoreductase; Monooxygenase; NADP; One-carbon metabolism; Iron.
 FT ACT_SITE 151 151 POTENTIAL.
 FT METAL 114 114 IRON 1 (BY SIMILARITY).
 FT METAL 144 144 IRON 1 (BY SIMILARITY).
 FT METAL 147 147 IRON 1 (BY SIMILARITY).
 FT METAL 209 209 IRON 2 (BY SIMILARITY).
 FT METAL 243 243 IRON 2 (BY SIMILARITY).
 FT METAL 246 246 IRON 2 (BY SIMILARITY).
 SQ SEQUENCE 525 AA; 59989 MW; FD5227DEEFAF9DDA CRC64;

Query Match 7.3%; Score 211; DB 1; Length 525;
 Best Local Similarity 21.0%; Pred. No. 5.2e-10;
 Matches 96; Conservative 71; Mismatches 196; Indels 94; Gaps 19;
 QY 45 KTHDWDKWDPPRLTMDAYWYQGEKELLY-AVIDAFTQNNNA----- 86
 DB 49 KYHMANETKEQFKVIAKEYARMEAAKDERQFGTLLDGLTRLGAGNKVHPRWGETMKVLSN 108
 QY 87 FLCVSDARYINALKFLQGVTPLE-----YLAHRGFAHVGRHFTGEGARIACQMSIDEIR 142
 DB 109 FLEVGEYNAIAASAMLDWDSATAAEQKNGYLA-----QVLDEIR 146


```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02120; AAA42785.1; -.
DR PIR; A03960; GNLJGB.
DR HSSP; P04585; IRT1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00075; rnaaseh; 1.
DR PFAM; PF00665; rve; 1.
DR PFAM; PF00078; rvt; 1.
KW Transferrase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 852 AA; 95182 MW; A7A63524A18C458D CRC64;

Query Match 3.7%; Score 107.5; DB 1; Length 852;
Best Local Similarity 19.0%; Pred. No. 0.35;
Matches 115; Conservative 70; Mismatches 214; Indels 207; Gaps 30;

QY 48 DWDKWDVDFRLTM-----DAYKYQGEKEK-----LYAVIDAFTQNNAFGLVSDARYI 96
Db 229 -----SPISLHQLQVLGDLQWVSRGPTTRPPLQLLYSSLKRRHDPRAITQLSPEQ-- 281
QY 97 NALKFLQGVPL-EYLHRG-----FAHVRHFTGEGARIAC-OMGOSIDELRY 144
Db 282 -----QGIAELROALSHNARSRYNEQEPFLAYV--HLTRAGSTLVLFQKGAQFPAYF 332
QY 145 QT-----ETHAMSTYKFFNGFHS-----NOWFD----- 169
Db 333 QTLPLDNOASPWGLLLGCGYLOQAALSSYAKTILKYHNPDKSLDNWQSSDPRVQ 392
QY 170 ---RWYLSVPKSPFEDAYSSGFFE-----FLTAVSFSEFYVLTNLLFVPMGSAAY 218
Db 393 ELLQLW-----POISSOGIQPPGPKWTLITRAEVFLTP-QFSDPTIPAALCL--FSDCAWG 445
QY 219 NG-----DMSTVTFGSAQSDSHMTLGTICIKFLLEQDDPDNPVIVQRWID-KW 267
Db 446 RGAYCLWKDHLDFQAVPAPESAQKGLAGLAA-----APPEPLMI-----WVDSKY 496
QY 268 FWRGYRLTLVAMMDYMQPKVMSWR----- 294
Db 497 LYSLLRTLVLGA-----WLQPDVPVSYALLYKSLLRHPALVGHVSRSHSASHPIASLNNY 552
QY 295 -----PSWEMYAQNGGALFK-----DLARYGIREPKGWQDACE-----GKDH 333
Db 553 VDQLPLETPEQWHKLTCHNSRALSRWPNRISANDPRSPATLCETCQKLNITGGCKMET 612
QY 334 SHOAWS-----TFYGFNAAS-AFHTWPTDEMGLWSAKYDPSDFRYRPRFD--- 380
Db 613 TORGWAPNHIWQADITHYKYQFTYALHVFVDYSGATHASAKRGLTTQTIEGLELAIV 672
QY 381 HWGEQAR-----AGNRFYMKTLPMLCQTCQIPMLFTEPGNPKIGARES-----NVLGNKF 431
Db 673 HLGREPKLNTDQGANYSKIFVRFCQFGISLSHVHPNTPSSGLVVERTNGLLKLLSLY 732
QY 432 HFCSDH-----CKDIFHEFPQKYV-----QAW-----LPVHQIHQGNCFPPDAD----- 470
Db 733 HLDEPHLPMQALSRAALWTHQINLPLKTRWELHSPPLAVISEGGETPKGSKFLFY 792
QY 471 --PGNEG--FDPLAALVDYAVTMGRON-----LDFGSEDEQKN 505
Db 793 KLPQGNRRRLGPLPALVEASGALLATNPVVPWRLLKAFKCLKNDGPDAPN 847

RESULT 7
POL_BLVJ
ID POL_BLVJ STANDARD; PRT; 852 AA.
AC P03361;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Bovine leukemia virus (Japanese isolate BLV-1) (BLV).
OC Viruses; Retroviral viruses; Retroviridae; BLV-HIV retroviruses.
KN [1]
RA Ikawa Y.;
RA Sagata N., Yasunaga T., Tsuzuku-Kawamura J., Ohishi K., Ogawa Y.,
RT "Complete nucleotide sequence of the genome of bovine leukemia virus:
RT its evolutionary relationship to other retroviruses.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:677-681(1985).
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC CC -!- SIMILARITY: VERY STRONG, WITH BOVINE LEUKEMIA VIRUS ENV
CC POLYPROTEIN FROM OTHER ISOLATES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02120; AAA42785.1; -.
DR PIR; A03960; GNLJGB.
DR HSSP; P04585; IRT1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00075; rnaaseh; 1.
DR PFAM; PF00665; rve; 1.
DR PFAM; PF00078; rvt; 1.
KW Transferrase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 852 AA; 95182 MW; A7A63524A18C458D CRC64;

Query Match 3.7%; Score 107.5; DB 1; Length 852;
Best Local Similarity 19.0%; Pred. No. 0.35;
Matches 115; Conservative 70; Mismatches 214; Indels 207; Gaps 30;

QY 3 TPTLKKKLKLDRYAMTRGLGNET-----TYQPM-----DKVFPYDRYEIGIKH 47
Db 169 SPTEQORSQCYQALAAERDLGQVASEKTSQTPSPVFLGQMVHQIYVYOSLPTQIS 228
QY 48 DWDKWDVDFRLTM-----DAYKYQGEKEK-----LYAVIDAFTQNNAFGLVSDARYI 96
Db 229 -----SPISLHQLQVLGDLQWVSRGPTTRPPLQLLYSSLKRRHDPRAITQLSPEQ-- 280
QY 97 NALKFLQGVPL-EYLHRG-----FAHVRHFTGEGARIAC-OMGOSIDELRY 144
Db 281 -----QGIAELROALSHNARSRYNEQEPFLAYV--HLTRAGSTLVLFQKGAQFPAYF 332
QY 145 QT-----ETHAMSTYKFFNGFHS-----NOWFD----- 169
Db 333 QTLPLDNOASPWGLLLGCGYLOQAALSSYAKTILKYHNPDKSLDNWQSSDPRVQ 392
QY 170 ---RWYLSVPKSPFEDAYSSGFFE-----FLTAVSFSEFYVLTNLLFVPMGSAAY 218
Db 393 ELLQLW-----POISSOGIQPPGPKWTLITRAEVFLTP-QFSDPTIPAALCL--FSDCAWG 445
QY 219 NG-----DMSTVTFGSAQSDSHMTLGTICIKFLLEQDDPDNPVIVQRWID-KW 267
Db 446 RGAYCLWKDHLDFQAVPAPESAQKGLAGLAA-----APPEPLMI-----WVDSKY 496
QY 268 FWRGYRLTLVAMMDYMQPKVMSWR----- 294
Db 497 LYSLLRTLVLGA-----WLQPDVPVSYALLYKSLLRHPALVGHVSRSHSASHPIASLNNY 552
QY 295 -----PSWEMYAQNGGALFK-----DLARYGIREPKGWQDACE-----GKDH 333
Db 553 VDQLPLETPEQWHKLTCHNSRALSRWPNRISANDPRSPATLCETCQKLNITGGCKMET 612
QY 334 SHOAWS-----TFYGFNAAS-AFHTWPTDEMGLWSAKYDPSDFRYRPRFD--- 380
Db 613 TORGWAPNHIWQADITHYKYQFTYALHVFVDYSGATHASAKRGLTTQTIEGLELAIV 672
QY 381 HWGEQAR-----AGNRFYMKTLPMLCQTCQIPMLFTEPGNPKIGARES-----NVLGNKF 431
Db 673 HLGREPKLNTDQGANYSKIFVRFCQFGISLSHVHPNTPSSGLVVERTNGLLKLLSLY 732
QY 432 HFCSDH-----CKDIFHEFPQKYV-----QAW-----LPVHQIHQGNCFPPDAD----- 470
Db 733 HLDEPHLPMQALSRAALWTHQINLPLKTRWELHSPPLAVISEGGETPKGSKFLFY 792
QY 471 --PGNEG--FDPLAALVDYAVTMGRON-----LDFGSEDEQKN 505
Db 793 KLPQGNRRRLGPLPALVEASGALLATNPVVPWRLLKAFKCLKNDGPDAPN 847

RESULT 8
AMY_BACST
ID AMY_BACST STANDARD; PRT; 549 AA.
AC P06279; Q45519;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

```

DE ALPHA-AMYLASE PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-GLUCAN
DN GLUCANOHYDROLASE).
GN AMYS.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-39.
RX MEDLINE; 85234394.
RA Nakajima K., Imanaka T., Aiba S.;
RT "Nucleotide sequence of the Bacillus stearothermophilus alpha-amylase
RT gene.";
RL J. Bacteriol. 163:401-406(1985).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=DY5/PHI300;
RX MEDLINE; 86008166.
RA Ihara H., Sasaki T., Tsuboi A., Yamagata H., Tsukagoshi N., Uda S.;
RT "Complete nucleotide sequence of a thermophilic alpha-amylase gene:
RT homology between prokaryotic and eukaryotic alpha-amylases at the
RT active sites.";
RL J. Biochem. 98:95-103(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NZ-3;
RX MEDLINE; 86195857.
RA Gray G.L., Mainzer S.E., Rey M.W., Lamsa M.H., Kindie K.L.,
RA Carmona C., Requaet C.;
RT "Structural genes encoding the thermophilic alpha-amylases of
RT Bacillus stearothermophilus and Bacillus licheniformis.";
RL J. Bacteriol. 166:635-643(1986).
RN [4]
RP SEQUENCE FROM N.A.
RA Suominen I., Karp M., Lautamo J., Knowles J., Mantsaelae P.;
RT "Thermotable alpha amylase of Bacillus stearothermophilus: cloning,
RT expression, and secretion by Escherichia coli.";
RL (In) Chaloupka J., Krumphanz V. (eds.);
RL Extracellular enzymes of microorganisms, pp.129-137, Plenum Press,
RL New York (1987).
RN [5]
RP SEQUENCE OF 1-122 FROM N.A., AND SEQUENCE OF 35-48.
RC STRAIN=DY-5;
RX MEDLINE; 86059211.
RA Tsukagoshi N., Iritani S., Sasaki T., Takemura T., Ihara H.,
RA Idota Y., Yamagata H., Uda S.;
RT "Efficient synthesis and secretion of a thermophilic alpha-amylase by
RT protein-producing Bacillus brevis 47 carrying the Bacillus
RT stearothermophilus amylase gene.";
RL J. Bacteriol. 164:1182-1187(1985).
CC [1]- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC [1]- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11450; AAA22235.1; -;
DR EMBL; X02769; CAA26547.1; -;
DR EMBL; M57457; AAA22227.1; -;
DR EMBL; M13255; AAA22241.1; -;
DR PIR; A00845; ALHSP.
DR PIR; A24549; A24549.
DR HSP; P06278; 1BPL.
DR PRAM; PF00128; alpha-amylase; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal.
FT SIGNAL 1 34
FT CHAIN 35 549 ALPHA-AMYLASE.

FT	ACT_SITE	268	268	BY SIMILARITY.
FT	ACT_SITE	272	272	BY SIMILARITY.
FT	ACT_SITE	365	365	BY SIMILARITY.
FT	CONFLICT	13	13	M -> V (IN REF. 3).
FT	CONFLICT	19	19	L -> W (IN REF. 3).
FT	CONFLICT	23	23	L -> S (IN REF. 2 AND 3).
FT	CONFLICT	31	31	P -> H (IN REF. 2 AND 5).
FT	CONFLICT	107	107	A -> T (IN REF. 2 AND 3).
FT	CONFLICT	167	167	T -> I (IN REF. 4).
FT	CONFLICT	179	179	P -> N (IN REF. 3).
FT	CONFLICT	251	251	S -> N (IN REF. 2, 3 AND 4).
FT	CONFLICT	260	260	TNI -> RTL (IN REF. 4).
FT	CONFLICT	284	284	D -> Y (IN REF. 2, 3 AND 4).
FT	CONFLICT	312	312	M -> T (IN REF. 2 AND 3).
FT	CONFLICT	338	338	T -> A (IN REF. 2 AND 3).
FT	CONFLICT	342	342	R -> S (IN REF. 3).
FT	CONFLICT	346	346	T -> N (IN REF. 3).
FT	CONFLICT	376	376	V -> C (IN REF. 2 AND 3).
FT	CONFLICT	526	527	WS -> RP (IN REF. 2).
FT	CONFLICT	527	527	S -> P (IN REF. 2).
FT	CONFLICT	535	535	D -> G (IN REF. 2 AND 3).
SO	SEQUENCE	549 AA;	62670 MW;	3A2DD93A955E79D3 CRC64;

Query Match 3.6%; Score 104; DB 1; Length 549;
Best Local Similarity 18.3%; Pred. No. 0.39;
Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

Qy	21	RGLG--WETTYQPMKVPFYDREYGI-----KIHDWDKW-----VDPFRITMDAY 63
Db	213	RGIGKAWDEVDTEGNYDYLMDLDHPEVVTLSKSGKGYVNTTNDIGFRLDAVKH 272
Qy	64	WKYQGEKEKKLYAVIDAFTQNNAFI--CVSDARYINALKLFLOGVTPLEYLAHRCFAHVG 121
Db	273	IKF-----SFFPDWLSVRSQTKPLFTVG----EYWSY----- 302
Qy	122	RHFTGEGARIACQMOSIDELRHQYOTETHAMSTYKFFNGEHSNOWNQEDRVWLSVPKSEFF 181
Db	303	-----DINKLHNYIMKTN--GTMSLFDALEN-----KFY 330
Qy	182	EDAYSSGPFELTAVSFSEYVLNLLFVPFMSGAAAYNGDMSTVTFGSAOSDSRHMIL 241
Db	331	TASKSGGTFDMRT-----LMTNTLM-----KDQPTL----- 356
Qy	242	GIECKFLEODPNVPVQVQWIDKWF-----WRGYRL----- 275
Db	357	---AVTFVDNHDTEPGQALQSWDPFKPLAYAFILTRQEGYPCVFGYGYGIPQYNIPS 413
Qy	276	-----TLVA-----MMMDYMPKRVMSW--RE-----SWEM 299
Db	414	LKSKIDPLLIARDYAYGTQHDYLDHSDIIGTWREGVTEKPGSGIAALITDPPGSKWMY 473
Qy	300	YAEONGGALFKDLARYGIREPKWODACEGKDHSQWSTFYGFNAASAFHTWVP---T 356
Db	474	VGKHAGKVFVDLT--GNR-----SDTVT-----INSDGWGEF--KVGSGSV--SVWVPRKTT 520
Qy	357	EDEMGWLSAKYP--DSEDRYRPRPDHW 382
Db	521	VSTTAWSTTRPTWTEDEFVRWTEPRLVAV 548

RESULT 9
MLH1 ARATH STANDARD; PRT; 526 AA.
ID O49621; O22766;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MLO PROTEIN HOMOLOG 1 (ATMLO-H1).
GN MLO-H1 OR T10p11.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;

```

OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA Panstruga R.:
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. JOHNSONIA;
RA Kaplan N., Johnson D., Schutz K., Gnoj L., Hoffman J., Till S.,
RA de la Bastide M., Granat S., Hameed A., Gottesman T., Hasegawa A.,
RA Shohdy N., Parnell L., Dedhia N., Johnson A.F., Lodhi M.,
RA Martienssen R., Chen E.Y., Wilson K., McCombie W.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MLO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95352; CAB08605.1; -.
DR EMBL; AC002330; AAC78258.1; -.
KW Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 309 329 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 413 433 POTENTIAL.
FT TRANSMEM 61 61 POTENTIAL.
FT CONFLICT 206 339 MISSING (IN REF. 2).
FT CONFLICT 393 435 VTGFDSCMGQVRYIVPRIVGFIQVLCYSYSLPLIYIV
FT (IN REF. 2).
FT SQ -> GLHSSALQLQYTSALRHRITGNKLIFFKKLENIL
FT (IN REF. 2).
FT SEQUENCE 526 AA; 59126 MW; 8E63C276A71A9768 CRC64;
Query Match 3.6%; Score 103; DB 1; Length 526;
Best Local Similarity 22.0%; Pred. No. 0.45;
Matches 71; Conservative 43; Mismatches 130; Indels 78; Gaps 17;
QY 45 KIHQDKWVDFRLTMDAYKYQGE---KEKKLYAVIDAFTQNNAFGLVSDARYI-----96
Db 184 RIHQKKWEDSI-----ADEKFDPTALRKRRVTHVHNHAFIKHEH-FUGIGKSDVILGWT 237
QY 97 -NALKLFQGVTPLEYLAHR-GFAHVGRHFTGEGARACQMSIDELRHYYQTETHAMSTY 154
Db 238 QSFLKQFVDSYTKSDYVTLRIGF--IMTHCKG-NPKLNFHKYMMRALEDDEDFKQVVGISWY 294
QY 155 NKFF-----NGFHHSNWFDRVWLVSPKSFEDDAYSGPPEFLTAVSFSEYVLT 205
Db 295 LWIEVWIFLLINVGWH-----TYFWIAFTI--PFALLAVGTLKLEHVIA 336
QY 206 NLLVFPMSGAYNGDM-----STVTFGSQSDSRHMTLGTGIEKTFLEQDPNPVIVQ 261
Db 337 QLAHEVAEKHVAIGDLVVKPSDEHFWFSKPO-----IVLYLTHFLFNQAFETIAFF- 388
QY 262 RWIDKWFWRGVRLLTLVAMMDYMOPK-----RVMSWSRESWEMYA--EQNGG-----AL 308
Db 389 ----FWIWTVTGFSICMGQVRYIVPRLVIGVFIQVLCYSYSLPLIYIVSQSGSFKKAI 444
QY 309 FKDLARYGIREPKGWQDACEK 330
Db 445 FEENVQVGL---VGWAQVKQK 463
RESULT 10
MCRA_METH
MCRA_METHH STANDARD; PRT; 549 AA.
C27232; Q50493;
15-JUL-1998 (Rel. 36, Created)
15-JUL-1998 (Rel. 36, Last sequence update)
15-FEB-2000 (Rel. 39, Last annotation update)
METHYL-COENZYME M REDUCTASE I ALPHA SUBUNIT (EC 1.8.-.-) (MCR I
ALPHA).
MCRA OR MTHL1364.
MCRA Methanobacterium thermoautotrophicum.
OS Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
CC [1]
RN RP SEQUENCE FROM N.A.
RA STRAIN=DELTA H;
RX MEDLINE; 95014084.
RA Pihl T.D., Sharma S., Reeve J.N.;
RT *Growth phase-dependent transcription of the genes that encode the
RT two methyl coenzyme M reductase isoenzymes and N5-
RT methyltetrahydromethanopterin:coenzyme M methyltransferase in
RT Methanobacterium thermoautotrophicum delta H.*;
RL J. Bacteriol. 176:6384-6391(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=DELTA H;
RX MEDLINE; 98037514.
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwanji N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT *Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.*;
RL J. Bacteriol. 179:7135-7155(1997).
RN [3]
RP SEQUENCE OF 1-18.
RA STRAIN=DELTA H;
RX MEDLINE; 91099370.
RA Rospert S., Linder D., Ellermann J., Thauer R.K.;
RT *Two genetically distinct methyl-coenzyme M reductases in
RT Methanobacterium thermoautotrophicum strain Marburg and delta H.*;
RL Eur. J. Biochem. 194:871-877(1990).
CC -1- FUNCTION: THIS ENZYME COMPLEX CATALYZES THE FINAL STEP IN
CC METHANOGENESIS, WHICH IS THE TERMINAL STEP OF ANAEROBIC
CC DEGRADATION OF BIOMASS.
CC -1- CATALYTIC ACTIVITY: REDUCTION OF METHYL-COENZYME M (2-(METHYLTHIO)
CC ETHANESULFONIC ACID) WITH 7-MERCAPTOHEPTANOYLPHREONINE PHOSPHATE
CC TO METHANE AND AN HETERODISULFIDE (CH(3)-S-COM + H-S-HTP = CH(4) +
CC COM-S-S-HTP).
CC -1- COFACTOR: THE ENZYME COMPLEX BINDS TIGHTLY (BUT NOT COVALENTLY)
CC TO TWO MOLECULES OF COENZYME F430. F430 IS A YELLOW NICKEL
CC PORPHINOID.
CC -1- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO GAMMA CHAINS.
CC -1- DEVELOPMENTAL STAGE: THERE ARE TWO MCR COMPLEXES IN THIS BACTERIA.
CC MCR II IS EXPRESSED IN THE EARLY GROWTH PHASE. LATE GROWTH CELLS
CC CONTAINS MOSTLY MCR I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10036; AAA73445.1; -.
DR EMBL; AE000885; AAB85653.1; -.
KW Methanogenesis; Oxidoreductase; Methylation; Multigene family.
FT INIT_MET 0
FT MOD_RES 256 256 METHYLATION (BY SIMILARITY).
FT MOD_RES 269 269 METHYLATION (BY SIMILARITY).
FT CONFLICT 287 287 D -> E (IN REF. 1).

```

Science 278:1457-1462(1997).

-1- FUNCTION: THIS ENZYME COMPLEX CATALYZES THE FINAL STEP IN METHANOGENESIS, WHICH IS THE TERMINAL STEP OF ANAEROBIC DEGRADATION OF BIOMASS.

-1- CATALYTIC ACTIVITY: REDUCTION OF METHYL-COENZYME M (2-(METHYLTHIO) ETHANESULFONIC ACID) WITH 7-MERCAPTOHEPTANOYLTHREONINE PHOSPHATE TO METHANE AND AN HETERODISULFIDE (CH(3)-S-COM + H-S-HTP = CH(4) + COM-S-S-HTP).

-1- COFACTOR: THE ENZYME COMPLEX BINDS TIGHTLY (BUT NOT COVALENTLY) TO TWO MOLECULES OF COENZYME F430. F430 IS A YELLOW NICKEL PORPHINOID.

-1- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO GAMMA CHAINS.

-1- DEVELOPMENTAL STAGE: THERE ARE TWO MCR COMPLEXES IN THIS BACTERIA. MCR II IS EXPRESSED IN THE EARLY GROWTH PHASE. LATE GROWTH CELLS CONTAINS MOSTLY MCR I.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@sib-sib.ch.

EMBL; X07794; CAA30639.1; --
DR PIR; E28544; E28544.
DR PDB; 1MRO; 11-NOV-98.
KW Methanogenesis; Oxidoreductase; Methylation; Multigene family; 3D-structure.
FT INIT_MET 0 0 METHYLATION.
FT MOD_RES 256 256 METHYLATION.
FT MOD_RES 269 269 METHYLATION.
SQ SEQUENCE 549 AA; 60379 MW; 979ABB4B6CB8622B CRC64;

Query Match 3.4%; Score 99; DB 1; Length 549;
Best Local Similarity 19.5%; Pred. No. 1;
Matches 84; Conservative 56; Mismatches 182; Indels 108; Gaps 18;

QY 141 LRHYQTETH---AMSTYKFKFNGPHHSNQWFDVRVWYLSVPKSPFED-----AYSSGPF 190
DB 145 VQEHMVETHPALVADSTYVKVFTGNDDEIADLPDAFVIDINKFPEDQAEYLKAEVGDGIW 204
QY 191 EFL---TAVSFSEFYVLT---NLIFVPFMSGAAYN---GDMSTVTFGSAQSDSRHML 241
DB 205 QVVRIPITVSTCDGATTSWSAMQIGMSMISAYKQAAGEAATGDFAYAKHAIEVIMGT 264
QY 242 GIECIKFLLEDPPDNVPIVQWIDKWFWRGYRLTLTYAMMDYMQPKV-----MS 292
DB 265 YLPVRRARGENEPGGVVF-----GYLADICQSSRVNEDPVRVLDVVATGAML 313
QY 293 WRSEWMYAFONGA-----LFKDLARYG---IREPKGQDACEGKDHIHQ 336
DB 314 YDQIWL-LGYSMSGVGFTQYATAAYTDNILLDDFTYFCKEYVEDKYGLCEAPNNMDTLDV 372
QY 337 AWS-TTFYGFNAASAFHWVTEDEMG-----WLSAKYDPSDFRYRPR 378
DB 373 ATEVTFYGLQEYEPYALL--EDQFGGSRQAAVVAAGCGSTAFATGATGLSGWYLSM 430
QY 379 FDHWGEQARAGNRYMKTLPMLCQTCQIPMLFTEPGNPTKI-GARESNYLGKNTFHFCSDH 437
DB 431 YLHKEQHSRLG--FYGYDLODQCGASNVFSIRGDEGLPLERGNYPNYAMNVGH----- 483
QY 438 CKDIFDHPQKYQVQAWLPVHQIHQGNCFPPDADPGAEGFDFLAALDYAVTGRONLDF 497
DB 484 -----QGEYAGISQA-----PHAAKGDAFVFNPLV-----KIAFADNLFV 519
QY 498 DGSDDQKNFA 507
DB 520 DFTNVRGEFA 529

RESULT 12

QY 141 LRHYQTETH---AMSTYKFKFNGPHHSNQWFDVRVWYLSVPKSPFED-----AYSSGPF 190
DB 145 VQEHMVETHPALVADSTYVKVFTGNDDEIADLPDAFVIDINKFPEDQAEYLKAEVGDGIW 204
QY 191 EFL---TAVSFSEFYVLT---NLIFVPFMSGAAYN---GDMSTVTFGSAQSDSRHML 241
DB 205 QVVRIPITVSTCDGATTSWSAMQIGMSMISAYKQAAGEAATGDFAYAKHAIEVIMGT 264
QY 242 GIECIKFLLEDPPDNVPIVQWIDKWFWRGYRLTLTYAMMDYMQPKV-----MS 292
DB 265 YLPVRRARGENEPGGVVF-----GYLADICQSSRVNEDPVRVLDVVATGAML 313
QY 293 WRSEWMYAFONGA-----LFKDLARYG---IREPKGQDACEGKDHIHQ 336
DB 314 YDQIWL-LGYSMSGVGFTQYATAAYTDNILLDDFTYFCKEYVEDKYGLCEAPNNMDTLDV 372
QY 337 AWS-TTFYGFNAASAFHWVTEDEMG-----WLSAKYDPSDFRYRPR 378
DB 373 ATEVTFYGLQEYEPYALL--EDQFGGSRQAAVVAAGCGSTAFATGATGLSGWYLSM 430
QY 379 FDHWGEQARAGNRYMKTLPMLCQTCQIPMLFTEPGNPTKI-GARESNYLGKNTFHFCSDH 437
DB 431 YLHKEQHSRLG--FYGYDLODQCGASNVFSIRGDEGLPLERGNYPNYAMNVGH----- 483
QY 438 CKDIFDHPQKYQVQAWLPVHQIHQGNCFPPDADPGAEGFDFLAALDYAVTGRONLDF 497
DB 484 -----QGEYAGISQA-----PHAAKGDAFVFNPLV-----KIAFADNLFV 519
QY 498 DGSDDQKNFA 507
DB 520 DFTNVRGEFA 529

Query Match 3.4%; Score 99; DB 1; Length 549;
Best Local Similarity 19.5%; Pred. No. 1;
Matches 84; Conservative 56; Mismatches 182; Indels 108; Gaps 18;

QY 141 LRHYQTETH---AMSTYKFKFNGPHHSNQWFDVRVWYLSVPKSPFED-----AYSSGPF 190
DB 145 VQEHMVETHPALVADSTYVKVFTGNDDEIADLPDAFVIDINKFPEDQAEYLKAEVGDGIW 204
QY 191 EFL---TAVSFSEFYVLT---NLIFVPFMSGAAYN---GDMSTVTFGSAQSDSRHML 241
DB 205 QVVRIPITVSTCDGATTSWSAMQIGMSMISAYKQAAGEAATGDFAYAKHAIEVIMGT 264
QY 242 GIECIKFLLEDPPDNVPIVQWIDKWFWRGYRLTLTYAMMDYMQPKV-----MS 292
DB 265 YLPVRRARGENEPGGVVF-----GYLADICQSSRVNEDPVRVLDVVATGAML 313
QY 293 WRSEWMYAFONGA-----LFKDLARYG---IREPKGQDACEGKDHIHQ 336
DB 314 YDQIWL-LGYSMSGVGFTQYATAAYTDNILLDDFTYFCKEYVEDKYGLCEAPNNMDTLDV 372
QY 337 AWS-TTFYGFNAASAFHWVTEDEMG-----WLSAKYDPSDFRYRPR 378
DB 373 ATEVTFYGLQEYEPYALL--EDQFGGSRQAAVVAAGCGSTAFATGATGLSGWYLSM 430
QY 379 FDHWGEQARAGNRYMKTLPMLCQTCQIPMLFTEPGNPTKI-GARESNYLGKNTFHFCSDH 437
DB 431 YLHKEQHSRLG--FYGYDLODQCGASNVFSIRGDEGLPLERGNYPNYAMNVGH----- 483
QY 438 CKDIFDHPQKYQVQAWLPVHQIHQGNCFPPDADPGAEGFDFLAALDYAVTGRONLDF 497
DB 484 -----QGEYAGISQA-----PHAAKGDAFVFNPLV-----KIAFADNLFV 519
QY 498 DGSDDQKNFA 507
DB 520 DFTNVRGEFA 529

Query Match 3.4%; Score 99; DB 1; Length 549;
Best Local Similarity 19.5%; Pred. No. 1;
Matches 84; Conservative 56; Mismatches 182; Indels 108; Gaps 18;

QY 141 LRHYQTETH---AMSTYKFKFNGPHHSNQWFDVRVWYLSVPKSPFED-----AYSSGPF 190
DB 145 VQEHMVETHPALVADSTYVKVFTGNDDEIADLPDAFVIDINKFPEDQAEYLKAEVGDGIW 204
QY 191 EFL---TAVSFSEFYVLT---NLIFVPFMSGAAYN---GDMSTVTFGSAQSDSRHML 241
DB 205 QVVRIPITVSTCDGATTSWSAMQIGMSMISAYKQAAGEAATGDFAYAKHAIEVIMGT 264
QY 242 GIECIKFLLEDPPDNVPIVQWIDKWFWRGYRLTLTYAMMDYMQPKV-----MS 292
DB 265 YLPVRRARGENEPGGVVF-----GYLADICQSSRVNEDPVRVLDVVATGAML 313
QY 293 WRSEWMYAFONGA-----LFKDLARYG---IREPKGQDACEGKDHIHQ 336
DB 314 YDQIWL-LGYSMSGVGFTQYATAAYTDNILLDDFTYFCKEYVEDKYGLCEAPNNMDTLDV 372
QY 337 AWS-TTFYGFNAASAFHWVTEDEMG-----WLSAKYDPSDFRYRPR 378
DB 373 ATEVTFYGLQEYEPYALL--EDQFGGSRQAAVVAAGCGSTAFATGATGLSGWYLSM 430
QY 379 FDHWGEQARAGNRYMKTLPMLCQTCQIPMLFTEPGNPTKI-GARESNYLGKNTFHFCSDH 437
DB 431 YLHKEQHSRLG--FYGYDLODQCGASNVFSIRGDEGLPLERGNYPNYAMNVGH----- 483
QY 438 CKDIFDHPQKYQVQAWLPVHQIHQGNCFPPDADPGAEGFDFLAALDYAVTGRONLDF 497
DB 484 -----QGEYAGISQA-----PHAAKGDAFVFNPLV-----KIAFADNLFV 519
QY 498 DGSDDQKNFA 507
DB 520 DFTNVRGEFA 529

Query Match 3.4%; Score 99; DB 1; Length 549;
Best Local Similarity 19.5%; Pred. No. 1;
Matches 84; Conservative 56; Mismatches 182; Indels 108; Gaps 18;

QY 141 LRHYQTETH---AMSTYKFKFNGPHHSNQWFDVRVWYLSVPKSPFED-----AYSSGPF 190
DB 145 VQEHMVETHPALVADSTYVKVFTGNDDEIADLPDAFVIDINKFPEDQAEYLKAEVGDGIW 204
QY 191 EFL---TAVSFSEFYVLT---NLIFVPFMSGAAYN---GDMSTVTFGSAQSDSRHML 241
DB 205 QVVRIPITVSTCDGATTSWSAMQIGMSMISAYKQAAGEAATGDFAYAKHAIEVIMGT 264
QY 242 GIECIKFLLEDPPDNVPIVQWIDKWFWRGYRLTLTYAMMDYMQPKV-----MS 292
DB 265 YLPVRRARGENEPGGVVF-----GYLADICQSSRVNEDPVRVLDVVATGAML 313
QY 293 WRSEWMYAFONGA-----LFKDLARYG---IREPKGQDACEGKDHIHQ 336
DB 314 YDQIWL-LGYSMSGVGFTQYATAAYTDNILLDDFTYFCKEYVEDKYGLCEAPNNMDTLDV 372
QY 337 AWS-TTFYGFNAASAFHWVTEDEMG-----WLSAKYDPSDFRYRPR 378
DB 373 ATEVTFYGLQEYEPYALL--EDQFGGSRQAAVVAAGCGSTAFATGATGLSGWYLSM 430
QY 379 FDHWGEQARAGNRYMKTLPMLCQTCQIPMLFTEPGNPTKI-GARESNYLGKNTFHFCSDH 437
DB 431 YLHKEQHSRLG--FYGYDLODQCGASNVFSIRGDEGLPLERGNYPNYAMNVGH----- 483
QY 438 CKDIFDHPQKYQVQAWLPVHQIHQGNCFPPDADPGAEGFDFLAALDYAVTGRONLDF 497
DB 484 -----QGEYAGISQA-----PHAAKGDAFVFNPLV-----KIAFADNLFV 519
QY 498 DGSDDQKNFA 507
DB

RESULT 12

Alternative splicing.
 KW FT SIGNAL 1 29
 FT CHAIN 30 1306
 ANGIOTENSIN-CONVERTING ENZYME, SOMATIC
 ISOFORM.
 FT DOMAIN 30 1259
 FT TRANSMEM 1260 1276
 FT DOMAIN 1277 1306
 FT REPEAT 227 583
 FT REPEAT 825 1181
 FT METAL 390 390
 FT ACT_SITE 391 391
 FT METAL 394 394
 FT METAL 988 988
 FT ACT_SITE 989 989
 FT METAL 992 992
 FT CARBOHYD 38 38
 FT CARBOHYD 54 54
 FT CARBOHYD 74 74
 FT CARBOHYD 79 79
 FT CARBOHYD 111 111
 FT CARBOHYD 146 146
 FT CARBOHYD 160 160
 FT CARBOHYD 272 272
 FT CARBOHYD 318 318
 FT CARBOHYD 445 445
 FT CARBOHYD 509 509
 FT CARBOHYD 677 677
 FT CARBOHYD 695 695
 FT CARBOHYD 714 714
 FT CARBOHYD 760 760
 FT CARBOHYD 942 942
 FT CARBOHYD 1191 1191
 FT CONFLICT 35 35
 FT CONFLICT 42 42
 SQ SEQUENCE 1306 AA; 149714 MW; 1B33BCA7301A26AA CRC64;

Query Match 3.3%; Score 94.5; DB 1; Length 1306;
 Best Local Similarity 18.8%; Pred. No. 7.3;
 Matches 142; Conservative 80; Mismatches 214; Indels 321; Gaps 46;

Qy 1 MDPTLKKKGLKRYAMTNGLGWETYY---QPMKV--PPYDEYEGIKIHWDKW--- 52
 Db 428 VSTPEHLKHGILLDR---VTNDESINLLKMALEKIAFLPF---GYLVQDW--RWGVF 479
 Qy 53 ---VDPRLTMDAYW---KYGEKEKLYAVIDATONNAFLGVSDARYINALKLFQ 105
 Db 480 SGRTPPSRYNFD--WYLLTKYQGG---ICPVVTRN---ETHFDGAKFHVPN 523
 Qy 106 VTP-----LEYLAHRGFARVGRH-----FTGGGARIAQOMQ----- 136
 Db 524 VTPYTRYFVSFLQFPHEALCKEAGYEGPLHQCDIYRSTRKAGAKLRKVLQAGSSRPQ 583
 Qy 137 -----SIDE---LRHYQETHAMSTYKNKFFNGFHHSNWFDRVWLVSPKFFEDA 184
 Db 584 VLKDMVGDLADLQKLLFYQFVPTQWLQEQNQ-QNG--EVLGWPYQWHPPLP-----DN 635
 Qy 185 YSSGPFPELT-----AVSFSEFVLNLLFVPMSCA--AYNGDMSTVT----- 226
 Db 636 YPEG-IDLVDEAEASKFVEYDRISQVWVNEYAEANNYNNTNITETSKILLQKNMQIA 694
 Qy 227 ---FGSAGS-----DPSRHMTLGIE----- 244
 Db 695 NHTLYGTQARKFDVNLQNTTIKRIKKVODLERAAALPAQBLEEYNKILLDMETYSVA 754
 Qy 245 -----CIKFLLEQDDPNPIVORWDK--WFWRGYRLLTLVAMMDYMQPKRYM--- 291
 Db 755 TVCHPNGSLQ--LEPDLTNMATSRKYEDLLWAEWGRDKAGRAILOFY--PKYVELIN 810
 Qy 292 -----SWRESWEMYA-BQNGGALFKDL-----ARYGIREPKGWQDAC 327
 Db 811 QAARLNGYVDAGDSWRSMTFSPLEQDLERLFEQLPQPLYNLHAYVVRALHHYCAQ--- 867

QY 328 EGDHIS-----HQAWSTFYG-----FNAASAFHTWVPTD--EMGWLSAKY 367
 Db 868 ---HINLEGPAPALLGNMWAQTWSNIYDLVVPFAPSMDT---TEAMLKOGWTPRRM 920
 QY 368 PDSFORYR-----PRFDHWG---EQARAGNR-----FY----- 393
 Db 921 FKAEADDFTSLLGLLPVPEF--WNKSMLEKPTDGRREVVCHASAWDFYNGKDFRIKQCTTV 978
 QY 394 -----MKTPLMLCQTCQIPMLFTEPCGNPTKIGARES----- 424
 Db 979 NLEDLVAAHEMCHIQFYMQYKDLPLVAREGANP---GFHEALGDVLALSVTPKHLHSL 1035
 QY 425 NYLGNK-----FHFCSDHCK-DIFDHE--POKYVQAMLPVHQ 458
 Db 1036 NLLSEGGSDHIDNLFMKMALDKIAFIPFSYLVQWRVDFGSGITKENYQEWWSLR 1095
 QY 459 IHQGNCFPP-----DADFGAEGFDPLAAVLDYYAVT 489
 Db 1096 KYQGLC-PPVPRQTQGFDFGAKFHIPSSVPYTRYFVS 1131

RESULT 14
 AMPN_LACDL STANDARD; PRT; 842 AA.
 ID AC P37896;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE AMINOPEPTIDASE N (EC 3.4.11.2) (LYSYL AMINOPEPTIDASE) (LYS-AP)
 DE (ALANINE AMINOPEPTIDASE).
 GN PEPN.
 OS Lactobacillus delbrueckii (subsp. lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Lactobacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 7290;
 RX MEDLINE; 94039025.
 RA Klein J.R., Klein U., Schad M., Plapp R.;
 RT "Cloning, DNA sequence analysis and partial characterization of pepN,
 a lysyl aminopeptidase from Lactobacillus delbrueckii ssp. lactis
 DSM7290.";
 RL Eur. J. Biochem. 217:105-114(1993).
 CC -!- FUNCTION: AMINOPEPTIDASE WITH BROAD SUBSTRATE SPECIFICITY TO
 SEVERAL PEPTIDES. OPTIMUM ACTIVITY IS MEASURED AT PH 6.5-7.0 AND
 TEMPERATURES 54-55 DEGREES CELSIUS.
 CC -!- COFACTOR: BINDS ONE ZINC ION.
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. IT MAY BE SECRETED THROUGH
 AN UNKNOWN MECHANISM.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);
 ALSO KNOWN AS THE PEPN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z21701; CAA79805.1; -.
 CC PIR; S38364; S38364.
 DR PFAM; PF01433; Peptidase M1; 1.
 DR PRINTS; PR00756; ALADIPTASE.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc.
 FT INIT_MET 0 0
 FT METAL 287 287
 FT ACT_SITE 288 288
 FT METAL 291 291
 FT METAL 310 310
 FT ACT_SITE 374 374

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:24:16 ; Search time 82.1 Seconds
(without alignments)
388.947 Million cell updates

Title: US-09-430-029-5
 Perfect score: 2873
 Sequence: 1 MDPTTLKKLGLKDRYAAMT.....FDGSEDKQNFAAWRGQATRN 516
 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database :      PIR_64:**
1:  pir1:**
2:  pir2:**
3:  pir3:**
4:  pir4:**
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1854.5	64.5	516	1	S44306	phenol 2-monoxyge
2	1846.5	64.3	517	1	D37831	phenol 2-monoxyge
3	1795	62.5	511	1	S47290	phenol 2-monoxyge
4	254.5	8.9	527	2	JQ0702	methane monooxygen
5	211	7.3	525	2	S15207	methane monooxygen
6	115	4.0	852	1	GNIJGA	pol polyprotein - bovi
7	111.5	3.9	852	2	S29358	pol polyprotein - bovi
8	107.5	3.7	852	1	GNIJGB	pol polyprotein - bovi
9	104	3.6	549	1	A24436	alpha-amylase (EC
10	103	3.6	526	2	T01089	hypothetical prote
11	103	3.6	550	2	B63022	methyl coenzyme M
12	99.5	3.5	1032	2	S74487	hypothetical prote
13	99	3.4	550	1	E28544	methyl coenzyme M
14	97	3.4	785	2	H72228	hypothetical prote
15	96.5	3.4	1209	2	T31657	reverse transcript
16	95	3.3	782	1	G64157	probable organic s
17	95	3.3	1336	3	T17479	hypothetical prote
18	95	3.3	1455	1	A48925	mannose receptor p
19	94.5	3.3	1306	1	A31759	peptidyl-dipeptida
20	94	3.3	395	2	T45547	arylsulfatase acti
21	94	3.3	452	2	T28094	hypothetical prote
22	93.5	3.3	333	1	S47288	phenol 2-monoxyge
23	93.5	3.3	579	2	JW0071	asparagine synthas
24	92.5	3.2	843	2	S38364	membrane alanyl am
25	91.5	3.2	548	1	ALBSF	alpha-amylase (EC
26	90.5	3.2	362	2	H69785	mannan endo-1,4-be
27	90.5	3.2	1256	1	A43829	muramidase-release
28	90	3.1	549	1	A24549	alpha-amylase (EC
29	89.5	3.1	1258	2	JC5765	inositol polyphosp

30	89	3.1	245	2	T23844	hypothetical protein
31	89	3.1	405	2	A35159	probable regulator
32	89	3.1	671	2	A45730	phenol 2-monooxygenase
33	89	3.1	722	2	B75074	methionyl-tRNA synthetase
34	89	3.1	875	2	A36369	1-phosphatidylinositol 3-kinase
35	88.5	3.1	848	2	T16430	hypothetical protein
36	88	3.1	549	1	A54541	alpha-amyrase (EC 3.2.1.1)
37	88	3.1	583	1	AJPMN2	asparagine synthase
38	88	3.1	713	2	B64743	lysine decarboxylase
39	87.5	3.0	561	2	S73087	alpha-amyrase (EC 3.2.1.1)
40	87.5	3.0	702	2	A34434	arylophorin alpha chain
41	87.5	3.0	1009	2	T18533	CryIAC toxin-binding protein
42	87.5	3.0	1210	3	T41740	probable cell wall protein
43	87.5	3.0	2410	2	T43435	alpha-glucan synthase
44	87.5	3.0	2410	2	T43731	probable cell wall protein
45	87.5	3.0	3119	2	T18414	protein g377 - like

ALIGNMENTS

RESULT 1
phenol 2-monooxygenase (EC 1.14.13.7) component N - Pseudomonas putida
S44306
N:Alternate names: phenolhydroxylase chain D
C:Species: Pseudomonas putida
C:Date: 10-Sep-1999 #sequence.revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: E58972; S54765; S44306; S47417
R:Ng, L.C.; Shindler, V.; Sze, C.C.; Poh, C.L.
Gene 151, 29-36, 1994
A:Title: Cloning and sequences of the first eight genes of the chromosomally encoded
A:Reference number: A58972; MUID:95129877
A:Accession: E58972
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-516 <NGL>
A:Cross-references: EMBL:X79063; NID:g483477; PIDN:CAA55663.1; PTD:g483481
A:Experimental source: strain P35X (NCBI 9869)
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994
R:Herrmann, H.; Mueller, C.; Schmidt, I.; Mahnke, J.; Petruschka, L.; Hahnke, K.
Mol. Gen. Genet. 247, 240-246, 1995
A:Title: Localization and organization of phenol degradation genes of Pseudomonas put
A:Reference number: S54761; MUID:95272534
A:Accession: S54765
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-63,'H',65-100,'F',102-106,'M',108-137,'E',140-512,'Q',514-516 <HE2>
A:Cross-references: EMBL:X80765; NID:g527546; PIDN:CAA56743.1; PID:g527550
A:Experimental source: strain H
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Genetics:
A:Gene: phlD; phnN
C:Superfamily: phenol 2-monooxygenase component N
C:Keywords: oxidoreductase

Query Match	64.5%;	Score 1854.5;	DB 1;	Length 516;
Best Local Similarity	64.8%;	Pred. No. 1.1e-142;		
Matches 329; Conservative	65;	Mismatches 105;	Indels 9;	Gaps 1;

QY	5	TLKKLGLKDRYAA	R	T	G	L	E	T	T	Y	O	P	M	D	K	V	F	P	D	R	E	G	I	K	H	D	K	W	D	P	F	L	T	M	D	A	Y	W	64									
Db	3	TNKKRLNLKDKY	R	L	T	D	L	G	W	E	P	S	Y	Q	K	K	E	D	V	F	L	E	H	F	E	G	I	K	T	D	W	K	E	D	P	F	L	T	M	D	S	Y	62					
QY	65	KYQGEKEKKIYA	N	I	D	A	F	T	N	N	A	F	L	G	S	D	A	R	I	N	A	L	K	F	L	G	O	V	T	P	L	E	V	I	A	H	R	G	F	A	H	V	R	H	F	124		
Db	63	KYQAEKEKKIYA	I	F	D	A	P	A	Q	N	G	H	N	I	S	D	A	R	V	N	A	L	K	F	L	T	G	S	P	L	E	Y	Q	A	F	G	S	R	V	R	G	R	F	122				
QY	125	TGEGARTACOM	S	T	D	E	L	R	H	C	T	T	H	A	M	S	T	N	K	T	F	E	N	G	F	H	S	N	O	F	E	D	R	V	W	I	L	S	V	P	K	S	F	E	D	A	184	
Db	123	SGAGARVACOM	A	I	D	D	V	R	H	V	O	T	Q	V	H	A	M	S	H	N	K	H	F	D	G	L	H	E	F	A	H	V	D	R	V	W	I	L	S	V	P	K	S	F	E	D	A	182
QY	185	YSSGPFPEFLTA	V	S	F	S	F	E	V	I	T	N	I	L	F	V	P	M	S	G	A	Y	N	G	D	M	S	T	V	T	E	G	S	A	O	S	E	S	R	H	M	I	L	G	I	E	A	244

Db 183 RTAGPEEFLTAVSPSEFYVLTNLLFVPFMSGAAYNGDMATVTFGSAQSDARHMTLGLE 242
QY 245 CIKELLQDDNPVIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWRSEWYAEQN 304
Db 243 VIKFMLEQHEDNVPITIORWIDKFWRGYRLTLTGMMDYMLPNKVMWSWSEAWGYVPEQA 302
QY 305 GGALFKDLARYGIREPKGWQDACGSKDHISHOAMSTFYGFNAASAFHTWVPTDEMGWLS 364
Db 303 GGALFKDLARYGIRPPKYVQTTIGKEHITHQVWGAFYQYSKATNFHTWIPGDEELNWL 362
QY 365 AKYPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTEPPGNPTKIGARES 424
Db 363 EKYEDTFDKYRPRFEWREQAQKGERFYNDTLPHLCQVCPALFTEPDDPTKLSRLSL 422
QY 425 NYLGNKFHSCDKDIFDHEPQKYQAWLPVHQIHGNCPPPPDADPGAGFDPPLAAVLID 484
Db 423 VHEGERYHFCSDGCDIFKNPEPKYIOAWLPVHQIYQGC-----EGGDVETVYQK 473
QY 485 YYAVTMGRDNLDFDGSDDOKNFAAWRGQ 512
Db 474 YYHKSGVDNLEYLGSPEHQRLALKGQ 501

RESULT 2

phenol 2-monooxygenase (EC 1.14.13.7) chain P3 - Pseudomonas sp. (strain CF600)

C:Species: Pseudomonas sp.

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: D37831

R:Nordlund, I.; Powlowski, J.; Shingler, V.

J. Bacteriol. 172, 6826-6833, 1990

A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phenol

A:Reference number: A37831; MUID:91072230

A:Accession: D37831

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-517 <NOR>

A:Cross-references: GB:M60276; GB:M37764; NID:gl51449; PID:AAA25942.1; PID:gl51453

C:Superfamily: phenol 2-monooxygenase component N

C:Keywords: oxidoreductase

Query Match 64.3%; Score 1846.5; DB 1; Length 517;
Best Local Similarity 64.8%; Pred. No. 4.8e-142;
Matches 328; Conservative 63; Mismatches 106; Indels 9; Gaps 1;

QY 7 KKKLGLKDRYAAMTGLGWETTYQPMQKVPYDREYEGIKIHWDKWDVPPRLTMDAYWKY 66
Db 6 KKRNLKDKRYLRDLAWETTYOKKEDVPLEHFEFEGIKITDWDKWDVPPRLTMDTYWKY 65
QY 67 QGEKKLYAVIDAFTONNAFLGVSDARYNALKFLQGVTPLEYLAHRGFARHFTG 126
Db 66 QAEKKLYAIFDAFQONNGHNSDARYNALKFLTAVSPLEYQAFQGSFVRGROFSG 125
QY 127 EGARIACOMQSIDELRHQYOTETHAMSTYKFNFGFHSNQWDFRVWYLSVPKSFEDAYS 186
Db 126 AGARVACOMQAIDELRHVQTVHAMSHYKHFQGLHDFAHMDYDRVWYLSVPKSYMDART 185
QY 187 SGPEFTTAVSFEYVLTNLLFVPFMSGAAYNGDMSTVTFGSAQSDSHRMTLGIECI 246
Db 186 AGPEFTTAVSFEYVLTNLLFVPFMSGAAYNGDMATVTFGSAQSDARHMTLGLEVI 245
QY 247 KFLLEDQDNPVIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWSEWYAEQNGG 306
Db 246 KFMLEQHEDNVPITIORWIDKFWRGYRLTLIGMMDYMLPNKVMSEAWGYVFEQAGG 305
QY 307 ALFKDLARYGIREPKGWQDACGSKDHISHOAMSTFYGFNAASAFHTWVPTDEMGWLSAK 366
Db 306 ALFKDLARYGIRPPKYVEQTTIGKEHITHQVWGALYQYSKATSEHTWIPGDEELNWLSEK 365
QY 367 YPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTEPGNPTKIGARES 426
Db 366 YPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTEPGNPTKIGARES 426

Db 366 YPDTFDKYRPRFEWREQAQKGERFYNDTLPHLCQVQQLPVITFTEPDDPTKLSRLSLVH 425
QY 427 LGNKFHSCDKDIFDHEPQKYQAWLPVHQIHGNCPPPPDADPGAGFDPPLAAVLIDY 486
Db 426 EGERYHFCSDGCDIFKNPEPKYIOAWLPVHQIYQGC-----EGGDVETVYQYY 476
QY 487 AVTMGRDNLDFDGSDDOKNFAAWRGQ 512
Db 477 HIKSGVDNLEYLGSPEHQRLALKGQ 502

RESULT 3

S47290

phenol 2-monooxygenase (EC 1.14.13.7) chain mopN - Acinetobacter calcoaceticus

N:Alternate names: phenol hydroxylase

C:Species: Acinetobacter calcoaceticus

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S70083; S47290

R:Ehrt, S.; Schirmer, F.; Hillen, W.

Mol. Microbiol. 18, 13-20, 1995

A:Title: Genetic organization, nucleotide sequence and regulation of expression of ge

A:Reference number: S70080; MUID:96154937

A:Accession: S70083

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-511 <EH2>

A:Cross-references: EMBL:Z36909; NID:9535279; PIDN:CAA95383.1; PID:9535283

A:Experimental source: strain NCIB8250

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 199

C:Genetics:

A:Gene: mopN

C:Superfamily: phenol 2-monooxygenase component N

C:Keywords: aromatic hydrocarbon catabolism; oxidoreductase

Query Match 62.5%; Score 1795; DB 1; Length 511;
Best Local Similarity 64.1%; Pred. No. 7.1e-138;
Matches 325; Conservative 60; Mismatches 112; Indels 10; Gaps 2;

QY 8 KKLGLKDRYAAMTGLGWETTYQPMQKVPYDREYEGIKIHWDKWDVPPRLTMDAYWKYQ 67
Db 13 KKLNAKERYRLTDLDDWDFSYADKDAFPYEEFEGIKITDWSKWDVPPRLTMDNYWKY 72
QY 68 GEKKLYAVIDAFTONNAFLGVSDARYNALKFLQGVTPLEYLAHRGFARHFTG 127
Db 73 AEKKLYAIFDAFQONNGHNSDARYNALKFLTAVTPLEYQAFQGSFVRGROFSGI 132
QY 128 GARIACOMQSIDELRHQYOTETHAMSTYKFNFGFHSNQWDFRVWYLSVPKSFEDAYS 187
Db 133 GARIASOMQSIDELRHVQTVHAMSHYKHFQGLHDFAHMDYDRVWYLSVPKSFEDAKSA 192

QY 188 GPFEFTTAVSFEYVLTNLLFVPFMSGAAYNGDMSTVTFGSAQSDSHRMTLGIECIK 247
Db 193 GPFEFTTAVSFEYVLTNLLFVPFMSGAAYNGDMATVTFGSAQSDARHMTLGLEIVK 252
QY 248 FLLLEDQDNPVIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWSEWYAEQNGGA 307
Db 253 FLLQEHEDNVPITIORWIDKFWRGYRLTLISIVGMMDYMLPNKVMSEAWGYVFEQAGG 312

QY 308 LFKDLARYGIREPKGWQDACGSKDHISHOAMSTFYGFNAASAFHTWVPTDEMGWLSAKY 367
Db 313 LFKDLARYGIRPPKYVEISKEKEHSHQAWWIFYNFHAAGFHTWIPTEMDWLSEKY 372
QY 368 PDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTEPGNPTKIGARES 426
Db 373 PDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTEPGNPTKIGARES 426

QY 427 LGNKFHSCDKDIFDHEPQKYQAWLPVHQIHGNCPPPPDADPGAGFDPPLAAVLIDY 486
Db 433 KDERYHFCSDGCDIFKNPEPKYIOAWLPVHQIYQGC-----EGGDVETVYQYY 483
QY 487 AVTMGRDNLDFDGSDDOKNFAAWRGQ 512

Db 484 NFNVGADNLDIEGSPDOORWKKWGNA 510

RESULT 4

QJ00702

methane monooxygenase (EC 1.14.13.25) component A alpha chain - Methylococcus capsulatus

C;Species: Methylococcus capsulatus

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999

C;Accession: JQ0702

R;Stainthorpe, A.C.; Lees, V.; Salmond, G.P.C.; Dalton, H.; Murrell, J.C.

Gene 91, 27-34, 1990

A;Title: The methane monooxygenase gene cluster of Methylococcus capsulatus (Bath).

A;Reference number: JQ0700; MUID:90382694

A;Accession: JQ0702

A;Molecule type: DNA

A;Residues: 1-527 <STA>

C;Comment: This multicomponent enzyme catalyzes the conversion of methane to methanol us

C;Genetics:

A;Gene: mmox

C;Keywords: oxidoreductase

Query Match 8.9%; Score 254.5; DB 2; Length 527;

Best Local Similarity 24.2%; Pred. No. 6.1e-13;

Matches 109; Conservative 65; Mismatches 179; Indels 97; Gaps 22;

QY 54 DPFLRTMDAYWKYQGEKEKKLYAVID-AFTONNAFLGVSDARYTNALKLFLQGVTPLEVL 112

Db 58 EOKFLIAKEYARMAVKDERQFSGISQVALTRLNAGRV-HFKWNETMKVV----- 106

QY 113 AHRGFVHGRH-----FTG---EGARIAQ-----MQSIDELRHQTEHMTSTYNKFF-- 158

Db 107 --SNFLEGEYNATAATGMLWDSQAABQKNGYLAQVLDEIRH-----THOCAYNYNFEAK 160

QY 159 NG---FHHSNQWFR---VWVLSVPKSFEDAYSSGFFELTAVSPF-----EYVLT 205

Db 161 NGQDPAGHNDARRRTTGPLW--KGMKRVSDGFISSG-----DAVECSLNLQLVGEACFT 213

QY 206 NLLFVPFMSGAAINGDMSTVTFGSAQSDSRHMTLGTIECLFLEQDPDNPVIVORWID 265

Db 214 NPLIVATERAANGDEIIPVFISIEDELHBMANGYQTV-VSIANDPASAKYINTDNL 272

QY 266 KFWRGVRLLT-LVAMMDYMPKRVMSWRESWEMIA-EQNGGALFKDLARYGIREPKGW 323

Db 273 NAFVTOQKIETPVLGLMFEYGSKEKVEPWKTNRVWYEDWGGIWRGLGYGVESPRSL 332

QY 324 QDACEGDHISHQAWSTFYGFNAASAFHTWVPTEDENGWLSAKYPDSFDRIYRFRFDHWG 383

Db 333 KADQDAYWAHHDYLLAYALWPTGFFRLALPDQEEWFEANYPGWYD-HYGKIYESEW- 390

QY 384 EQARAGNRFYMKTLPLMLCQTCQIP-----MLFTEPCGNPTKI-----CA- 421

Db 391 -RARG-----CEDSSGEFILPMWFIENNHPYIDRVSVQVPCPSLAKGAS 434

QY 422 --RESNYLGNKHFCHSDCKDIFDHEPOKY 449

Db 435 TLRVHEYNGQMHTEFSDQWGERMMLAEIPRY 464

RESULT 5

SI5207

methane monooxygenase (EC 1.14.13.25) component A alpha chain - Methylosinus trichospori

C;Species: Methylosinus trichosporium

C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 18-Jun-1999

C;Accession: SI5207; A39049

R;Cardy, D.I.N.; Laidler, V.; Salmond, G.P.C.; Murrell, J.C.

Mol. Microbiol. 5, 335-342, 1991

A;Title: Molecular analysis of the methane monooxygenase (MMO) gene cluster of Methylosi

A;Reference number: SI5207; MUID:91251762

A;Accession: SI5207

A;Molecule type: DNA

A;Residues: 1-525 <CAR>

A;Cross-references: EMBL:X55394; NID:g44613; PIDN:CAA39068.1; PID:g44614

R;FOX, B.G.; Liu, Y.; Dege, J.E.; Lipscomb, J.D.

J. Biol. Chem. 266, 540-550, 1991

A;Title: Complex formation between the protein components of methane monooxygenase fr

A;Reference number: A39049; MUID:91093180

A;Accession: A39049

A;Status: preliminary

A;Molecule type: protein

A;Residues: 2-16 <FOX>

C;Genetics:

A;Gene: mmox

C;Keywords: oxidoreductase

Query Match 7.3%; Score 211; DB 2; Length 525;

Best Local Similarity 21.0%; Pred. No. 2.1e-09;

Matches 96; Conservative 71; Mismatches 196; Indels 94; Gaps 19;

QY 45 KIHWDKWDVDFRLTMDAYWKYQGEKKLY-AVIDAFTQNNNA----- 86

Db 49 KYHMANETKEQFKVIAKEYARMEAAKDERQFGTLLDGLTRLGAGNKVHPRWGCTMKVTSN 108

QY 87 FLGVSADARYINALKLFLOGVTPLE-----YLAHRGFVHGRHFTGEGARIACOMQSIDEIR 142

Db 109 FLEVGEYNAAIAASAMLDWSATAAEQKNGYLA-----QVLDEIR 146

QY 143 HYQIETHAMSTYNKPFENGFIHS-----NQWFDVWYLSVPKSFEDAY-SSGPTEF 192

Db 147 H-----THOCARTINHYSKHYHDPAGHNDARRTRAGPLW--KGMKRVFADGFISSDAVE 200

QY 193 LTAVSFSEFYVLNLLFVPMFMSGAAINGDMSTVTFGSAQSDSRHMTLGTIECLFLEQ 252

Db 201 SVNQLQVGDTCFTNPLIVAVTEWALGNDEITPTVFLSVETDELHBMANGYQTV-VSIAN 259

QY 253 DPNVPIVQRIWIDKFWRGYRLLT-LVAMMDYMPKRVMSWRESWEMIAEQN-GGALFK 310

Db 260 DPASAKFLNTDLNNAFTQKQYFTFVLGLFELFYGSKEKVEPWKTNRVWYEDWGGIWRG 319

QY 311 DLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTW-----VPTEDMGW 362

Db 320 RLKYGIV-ESRVLRAKRDAYWAHHDL-----ALAAIAMWPLAFARLALPDEEDQAW 370

QY 363 LSAKYPSDFRYYRPRFDHWG---EQARAGNRFYMKTLIP-----MLQQTCTQIPMLFTEP 413

Db 371 FEANYPGWAD-HYGIKFNKWLKGYEDPKSGFIPYQWLLANGHDYVIDRVSVQVFI--P 426

QY 414 GNPKIGA-RSNYLGKHFCHSDCKDIFDHEPOKY 449

Db 427 SLAKGTSLRVHFEFGKKSLSLTDWGERQWLLEPERY 463

RESULT 6

GNLJGA

pol polyprotein - bovine leukemia virus (strain Australia)

N;Contains: endonuclease (EC 3.1.-.-); RNA-directed DNA polymerase (EC 2.7.7.49)

C;Species: bovine leukemia virus, BLV

A;Note: host Bos sp. (cattle)

C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999

C;Accession: JQ0555

R;Coultston, J.; Naif, H.; Brandon, R.; Kumar, S.; Khan, S.; Daniel, R.C.W.; Lavin, M.

J. Gen. Virol. 71, 1737-1746, 1990

A;Title: Molecular cloning and sequencing of an Australian isolate of proviral bovine

A;Reference number: JQ0554; MUID:90362060

A;Accession: JQ0555

A;Molecule type: DNA

A;Residues: 1-852 <COU>

A;Cross-references: DBJ:D00647; NID:g221051; PIDN:BAA00544.1; PID:g1000999; PID:g221

A;Note: This reading frame extends between two stop codons and does not begin with a

A;Note: the authors translated the codon CCC for residue 514 as Gln

C;Comment: The precise boundary between RNA-directed DNA polymerase and endonuclease

C;Genetics:

A;Gene: pol

C;Superfamily: pol polyprotein

C;Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; reverse tra

Query Match 4.0%; Score 115; DB 1; Length 852;
Best local similarity 18.7%; Pred. No. 0.23;
Matches 134; Conservative 81; Mismatches 252; Indels 248; Gaps 36;

```
QY      3 TPTLKKLGLKDRYAAMTRGLGWET-----IYQPM-----DKVFFYDRYEGIKIH 47
Db      169 SPTGEORLCQYQTAAHLRLDGLGVASEKTRQTPSPVFLQGMVHERMVTYQSILPTLOIS 228
QY      48 DWDKWDVPDFRLTM-----DAYWKYQGEKEKK-----LYAVIDAFTONNAFLGVSDARYI 96
Db      229 -----SPISLHQLQTVLGDLOWVSRGPIITRRPLQLLYSLKIGIDDPRAIIHLSPQOQ- 281
QY      97 NALKLFTQGVTP-L-EYLAHRG-----FAHVGRHFTGEGARIAC-QMOSIDELRHY 144
Db      282 -----QGIATLRLQALSHNARSRYNEQEPILAYV--HLTRAGSTLVLFQKGAQFPLAYF 332
QY      145 QT-----ETHAMSTYKNFENGFIHS-----NWFN----- 169
Db      333 QTPLTDNQASPDWGLLLLLGCOYLQAQLSSYAKTILKYHNLPKTSLDNWIOQSEDPVQ 392
QY      170 ---RWYLSVPKSPFEDAYSSGPE-----FLTAVSFSEYVVLNLLFVPMGAAY 218
Db      393 ELLQLW-----POISSQGTQPPGPKWKTIVTRAEEVLTLP-QFSPEPIPAALCL--FSDGAAR 445
QY      219 NG-----DMSTVTTFGSAQSDRESRMTLGIECIKFLLEQDPDNVPIVQWID-KW 267
Db      446 RGAYCLWKDHLDDFQAVPAPESAQKGLAGLAGLAA---APPEPLNI-----WVDSKY 496
QY      268 FWGRYRLITIVAMMDYMQKRVMSWR----- 294
Db      497 LYSLLRLTIVLGA-----WLQDDPVSYALLKSLRLHPAIFVGHVRSSSHASHPIASLNNY 552
QY      295 -----ESWPMYAEQNGALFK-----DLARYGIREPKGWODACE-----GKDHI 333
Db      553 VDQLLPLETPEQWKLTHCNSRALSRWPNRISAWDPRSPATLCETCQRLNPTGGKMWRT 612
QY      334 SHQAWS-----TFYGENAAS-AFTWVPTEDMGWLSAKYPDSFDRYRPRFD--- 380
Db      613 IORGWAPNHIIQWADITHYKYKQFTYALJHVFDTVSGATHASAKRGLTTQTITIEGLLEAIV 672
QY      381 HWGEOQAR---AGNRFYMKTLPMLCQPCQIPMLFTPEGNTPKIGARES-----NYLGNKF 431
Db      673 HLGPRKLTNDQGANVYTSKTFVRFCQFGISLSHHVYVNTSSGLVERTNGLKLLLSKY 732
QY      432 HFCSDH-----CKDIFDHEPQYV---QAW-----LPVHIHQNCFFPAD----- 470
Db      733 HLDSPHPLPMTQALSRALWTHNQINLPLILKTRWELHHSPLAVISEGGETPKGSKDLFLY 792
QY      471 --PGAEG---FDPLAAVLDYVAVTMGRON-----LDFDGSSEDQKN 505
Db      793 KLPQONNRFWLGPLPALVEASGGALLATNPVWVFWRLLLKAFKCLKNDGPEADPN 847

RESULT 7
S29358
pol protein - bovine leukemia virus
C:Species: bovine leukemia virus, BLV
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
C:Accession: S29358
R:Rice, N.R.; Stephens, R.M.; Burny, A.; Gilden, R.V.
Virology 142, 357-377, 1985
A:Title: the gag and pol genes of bovine leukemia virus: nucleotide sequence and analysis
A:Reference number: S29356; MUID:86045859
A:Accession: S29358
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-852 <R1C>
A:Cross-references: EMBL:M10987
C:Superfamily: pol polyprotein
```

Query Match 3.9%; Score 111.5; DB 2; Length 852;
Best Local Similarity 19.0%; Pred. No. 0.45;
Matches 115; Conservative 70; Mismatches 214; Indels 207; Gaps 30;

QY 3 TPTLKKLGKLDKRDVAAATPGLGWET-----TYQPM-----DKVEPYDREGIKH 47
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 169 SPTSEQRSCQYQAALARRDLGFVASEKTQTTPSPVFELGQMWHNQIVITQSPTQIS 228
QY 48 DWDKWVDPERLTM-----DAYKYQGEKEK-----LYAVIDAFTONNAFLGVPSDARYI 96
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 229 -----SPLSHLOQAVLGDLOWVRSGTGRPRPLQLLYSLIKGTDPRAIIQLSPE-- 280
QY 97 NALKLFLOGVTPL-EYLAHRG-----FAHVGRRHTCEGARIAC-QMOSIDELHY 144
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 281 -----LOGIAELKOALSHNARSRYNEOPELLAY--HLTRAGSTLVLFORGAOFFPLAYF 332
QY 145 QT-----ETHAMSTYNKFENGFHH-----NQWFD----- 169
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 333 CTPLTDNQASPWGLLLILGCQVLQTAALLSSYAKPKLYKYYHNLPKTSLDNWIOSSEDPRVQ 392
QY 170 ---RWYLSVPKSFEDAYSQPE-----FLTAVSFSEFVVLNLFPVPSMAAY 218
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 393 ELLQIW----POISSQGTPPGPKTLITRAEVFLTP-QFSPEPIPAALCL--FSDGATG 445
QY 219 NG-----DMSTVTFGSADESRRMTLTGECIKELLEODPNVPVQSWID-KW 267
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 446 KGAYCLKWDHLLDFOAVPAPESAQKGELAGLAGLAA---APPEPLNI-----WVDSKY 496
QY 268 FWRGYRLLTVAMMDYMOPKRVMWSR----- 294
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 497 LYSLLRTVLGA---WLQPPDPVPSYALLYKSLLRHHPAIFGVHRSHSSASHPIASLNNY 552
QY 295 -----ESWEYAEONGALFK-----DIARYGIREPKGWODACE-----GKDHI 333
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 553 VDQLPLETTPSQWKHLTHCSRALSRLNPNRISAWDPSPATLCETCQRLNPTGGGMKT 612
QY 334 SHQAWS-----TFYGFNAAAS-AFTWTWPTYTEDMGWLSAKYPDGFDRYRPF-- 380
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 613 LQRGWAPHNIWQADITHYKYQFYVALHVVDVITYSGATHASAKRGLITOMIEGLEAIV 672
QY 381 HWGEAR-----AGNRFYMKTLIPMLCQCQIIPMLTEPCNPTKIGARES-----NYLGNKF 431
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 673 HLRPKKLNTDQGANYTSKTEVRCQQFGISLSHHVPYNPPTSGLVERTNGLLKLLSKY 732
QY 432 HFCSDH 437
||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
DB 733 HLDEPH 738

RESULT 8
GNLJGB
pol polyprotein - bovine leukemia virus
N:Contains: endonuclease (EC 3.1.-.-); RNA-directed DNA polymerase (EC 2.7.7.49)
C:Species: bovine leukemia virus, BLV
A:Note: host Bos sp. (cattle)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 01-May-1998
C:Accession: A03960
R:Sagata, N.; Yasunaga, T.; Tsuzuku-Kawamura, J.; Ohishi, K.; Ogawa, Y.; Ikawa, Y.
Proc. Natl. Acad. Sci. U.S.A. 82, 677-681, 1985
A>Title: Complete nucleotide sequence of the genome of bovine leukemia virus: its evolution
A:Reference number: A94063; MUID:85140159
A:Accession: A03960
A:Molecule type: DNA
A:Residues: 1-852 <SAG>
C>Note: the authors translated the codon TTC for residue 104 as Ser and CTA for resid
C:Comment: Specific enzymatic cleavages may yield mature proteins including reverse t
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein
C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; reverse tra

Query Match 3.7%; Score 107.5; DB 1; Length 852;

Query Match
3.7%; Score 107.5; DB 1; Length 852;

Best Local Similarity 19.0%; Pred. No. 0.94;
Matches 115; Conservative 70; Mismatches 214; Indels 207; Gaps 30;

```
QY 3 TPTLKKGLKDRYAAATRGWGWEI-----TYQPM-----DKVFPDYRYEGIKTH 47
Db 169 SPTEDORSQYALAAARLDLGFQVASEKTSQPSVPFLGQWVHQIVTYQSLPTLQIS 228
QY 48 DWDKVVDPFLPM-----DAYWKYQGEKEK-----IYAVIDAFTQNNAFNGSDARYI 96
Db 229 -----SPISLHQAVLQWVGSRGTTTRPLQLLYSSLKRHHDPRAIIQLSEQ-- 280
QY 97 NALKLELQGVTEL-EYLARG-----FAHVGRRHTGCGARTAC-OMOSIDELRIY 144
Db 281 -----LOGIAELRQALSHNARSRYNEQEPFLAYV--HLTRAGSTLVLFQKGAQFPPLAYF 332
QY 145 QT-----ETHAMSTYNKFFNGFRHS-----NOMED----- 169
Db 333 QPFLTDNQASPWGILLLLGCQYLQALSSYAKPIKLYHNLPKISLDNWIOSSEDPVQ 392
QY 170 -----RWYLSVPKSFEDAYSSGPFE-----FLTAVSFSEFYVLTNLLFVPMFSGAAY 218
Db 393 ELLQLW-----POISSOGIQPPGPWKTLITRAEVELTP-QFSPDPIPAALCL--FSDGATG 445
QY 219 NG-----DMSTVTFEGSAQSDSRHMTLIGTICIKELLEDQDNDVPIVORWID-KW 267
Db 446 RGAYCLWKDHLDFQAVPAPEASQAGELAGLAG-----LAAAPPEPVNI---WYDSKY 496
QY 268 FWRGYRLTLVAMMDYMPKRVMSNR-----LAAAPPEPVNI---WYDSKY 496
Db 497 IYSLRLTLVGA-----WLQDPVPVSALLYKSLLRHPALVGVHRSASHPIASLNNY 552
QY 295 -----ESWEMVYAEONGALFK-----DLARYGIREPKGWQDACE-----GKDHI 333
Db 553 VDOLLPLETPEQWHKLTCHNSRALSRWPNRISAWDRSPATLCEFCQKLNPTGGCKMET 612
QY 334 SIQANS-----TFYGFNAAS-AFTWVTEDEMGWLSAKYDPSDFRYRPRFD--- 380
Db 613 IORGWAPNHIAQADITHYKYKFIYALHVFVDITYSGATHASAKRGLTQTOTIEGLEAIV 672
QY 381 HWGEQAR-----AGNRFYMTLPMLCQTCOIPMLFTEPGNPTKIGARES-----NYLGKRF 431
Db 673 HLCGRPKLNTDOGANYSKTIFRFQCGVSIHHVFPNPTSSGLDTERINGLLKLLLSKY 732
QY 432 HPCSDH 437
Db 733 HLDEPH 738
```

RESULT 9
A24436
alpha-amylase (EC 3.2.1.1) precursor - Bacillus stearothermophilus plasmid pat5
N:Alternate names: 1,4-alpha-D-glucan glucanohydrolase
C:Species: Bacillus stearothermophilus
C>Date: 05-Jun-1987 #sequence_revision 18-Aug-1995 #text_change 18-Jun-1999
C:Accession: A24436; 139777
R:Nakajima, R.; Imanaka, T.; Alba, S.
J. Bacteriol. 163, 401-405, 1985
A:Reference number: A24436; MUID:85234394
A:Accession: A24436
A:Molecule type: DNA
A:Cross-references: 1-349 <NAK>
A:Experimental source: GB:M11450
A:Note: amino end of the mature protein also determined
R:Jorgensen, P.L.; Hansen, C.K.; Poulsen, G.B.; Diderichsen, B.
Gene 96, 37-41, 1990
A:Title: In vivo genetic engineering: homologous recombination as a tool for plasmid con
A:Reference number: 139772; MUID:91092499
A:Accession: 139777
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <RES>
A:Cross-references: GB:M62638; NID:g142514; PIDN:AAA22242.1; PID:g142515

C:Comment: Alpha-amylase genes have been found on plasmids and in multiple copies on
C:Genetics:
A:Gene: amys
A:Genome: plasmid
A:Start codon: GTG
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucoisidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology
C:Keywords: extracellular protein; glycosidase; heat-stable protein; hydrolase; polys
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-549/Product: alpha-amylase #status experimental <MAT>
F:235-368/Domain: alpha-amylase core homology <AMY>
F:139,237,272/Binding site: calcium (Asp, Asp, His) #status predicted
F:268,298,365/Active site: Asp, Glu, Asp #status predicted

Query Match 3.6%; Score 104; DB 1; Length 549;
Best Local Similarity 18.3%; Pred. No. 1;
Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

```
QY 21 RGLG--WEITYQPMKVFYDRIYEGT-----KIHDWKW-----VDPELTMDAY 63
Db 213 RGIGKAWDEVPDENGNYDILMYADLMDHPFVVTELKSGWKYVNTNIDGFRLDVAKH 272
QY 64 WKYQGEKEKKLYAVIDAFTQNNAF--GVSDARYINALKLELQGVTPLEYLAHRGFAHV 121
Db 273 IKF-----SFFPDWLSVRSQTKPLFTVG-----EWSY----- 302
QY 122 RHTGEGARACOMQSIDELRHYQTEHAMSTYNKFFNGPHHSNQWDFRVWYLSVPKSF 181
Db 303 -----DINKLHNYIMKTN--GTMSLFDAPLN-----KFY 330
QY 182 EDAYSSCPPEFLTAVSFSEFYVLTNLLFVPMFSGAAYNGDMSTVTFGFSNQSDSRHML 241
Db 331 TASKSGGTDMRT-----LMNTLM-----KDQPTL----- 356
QY 242 GTICIKFLELQDNDVPIVORWIDKWF-----NRGYRL----- 275
Db 357 ---AVTFVDNHDTPEQALQSWDPWPKPLAYAFILTRQEGYPCVFGYGYGIPQYNIPS 413
QY 276 -----TLVA-----MMDYMPKRVMSW-RE-----SWEM 299
Db 414 LKSKIDPLLIARDYAYGTQHDYLDHSDIIGTWREGTEKPGSGLAALITDGGSKWMY 473
QY 300 YAPONGALFKDLARYGIREPKGWQDACEGKDHISHQAWSTFFYGFNRSASFHTWVP---T 356
Db 474 VGQKHAGKVFDLT--GNR-----SDTIVT-----INSDGWGEF-KVNGGSV-SYMWVPRKTT 520
QY 357 EDEMGWLSAKYP--DSFDRYRPRFDHW 382
Db 521 VSIANSITRPTWIDEFVWTEFLVAV 548
```

RESULT 10
T01089
hypothetical protein T10P11.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 14-May-1999
C:Accession: T01089
R:Kaplan, N.; Johnson, D.; Schutz, K.; Gnoj, L.; Hoffmann, J.; Till, S.; de la Bastide
hi, M.; Martienssen, R.; Chen, E.Y.; Wilson, R.; McCombie, W.R.
submitted to the EMBL Data Library, November 1998
A:Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.
A:Reference number: Z14248
A:Accession: T01089
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-526 <KAP>
A:Cross-references: EMBL:AC002330; NID:g2262135; PID:g3892049
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4

A;introns: 61/1; 145/3; 184/3; 205/1; 237/3; 262/3; 309/1; 322/3; 392/3; 415/2; 435/3
A;Note: T10P11.12

Query Match 3.6%; Score 103; DB 2; Length 526;
Best Local Similarity 22.0%; Pred. No. 1.2;
Matches 71; Conservative 43; Mismatches 130; Indels 78; Gaps 17;
QY 45 KIHWDKWDVDFRLTMDAYKYGQE---KEKKLYAVIDAFTONNAFLGVSDARYI----- 96
DB 184 RIHWKKWEDSI-----ADEKFDDETALRRRVTHVHNHAFTEKH-FIGICKDSVILGWT 237
QY 97 -NALKLFQGYPTPLEYLAHR-GFAHVGHRHFTGEGARIACQMSIDELRHYOTETHAMSTY 154
DB 238 QSELKQFVDSVTKSDYVTLRLGF--IHTHCKG-NPKLNFHXYMRALEDDEKQVGVISWY 294
QY 155 NKFF-----NGFHNSQWEDRVVWLSVPKSFEDDAYSGGFETAVTSFSEYVLT 205
DB 295 LWIEVFIPLLNNGWH-----TYFWIAFT--PFALLAVGTKLEHVIA 336
QY 206 NLLFVPMMSGAAVNGDM---STVTFGSAQSDSRHMTLGIECTKFLLEDOPNPVPIVQ 261
DB 337 QLAHEVAEKHVAIEDGLVWPKSDHFEKPKQ-----IVLYLHLFQNAFEIAFT- 388
QY 262 RWIDKFWRGYRLTLVAMMDYMQPK-----RVMSWRESWEMYA--EONGG-----AL 308
DB 389 ----FWIWTYVGFDSCTMGQVRYIVPRVIGVFIQVLCYSTPLPLAIYVQMSGSFKA 444
QY 309 FKDLARYGIREPKWQDACEGK 330
DB 445 FEENVQVGL---VGWAQVKQK 463
RESULT 11
B69022
methyl coenzyme M reductase (EC 1.8.-.-) I alpha chain - Methanobacterium thermoautotroph
C;Species: Methanobacterium thermoautotrophicum
A;Variety: strain Delta H
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jun-1999
C;Accession: B69022; S78579
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
K.; S. Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514
A;Accession: B69022
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-550 <MTH>
A;Cross-references: GB:AE000885; GB:AF000666; NID:g2622256; PIDN:AAB85653.1; PID:g262227
A;Experimental source: strain Delta H
R;Rospert, S.; Linder, D.; Ellermann, J.; Thauer, R.K.
Eur. J. Biochem. 194, 871-877, 1990
A;Title: Two genetically distinct methyl-coenzyme M reductases in Methanobacterium therm
A;Reference number: S13864; MUID:91099370
A;Accession: S78579
A;Molecule type: protein
A;Residues: 2,'X',4-15 <ROS>
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH164
C;Complex: heterohexamer of two alpha, two beta (see PIR:F69022) and two gamma (see PIR:
C;Function:
de and methane
A;Pathway: methanogenesis
C;Superfamily: methyl coenzyme M reductase alpha chain
C;Keywords: heterohexamer; metalloprotein; methanogenesis; nickel; oxidoreductase
F;2-550/Product: methyl coenzyme M reductase I alpha chain #status experimental <MAT>
F;147/Binding site: coenzyme F430 nickel (Gln) (axial ligand) #status predicted
F;333/Active site: Tyr #status predicted
F;444/Binding site: coenzyme M (Tyr) #status predicted

F;481/Binding site: coenzyme B (Asn) #status predicted

Query Match 3.6%; Score 103; DB 2; Length 550;
Best Local Similarity 19.5%; Pred. No. 1.3;
Matches 85; Conservative 54; Mismatches 178; Indels 118; Gaps 17;
QY 141 LRHYOTETH---AMSTYKFNFGPHHSNWFDRVWLSVPKSFED------AYSSGPF 190
DB 146 VOEHMVETHPALVADSYVYKVTGNDIEADIDPAEVIDINKOFFEDQAFETKAEVGDGIW 205
QY 191 EFL---TAVSFSEYVLT---NLLFVPMMSGAAVY---GDMSTVTFGSAQSDSRHMTL 241
DB 206 QVRIPTIVSRCTCGATTSRWSAQMGSMISAYKQAAGEAATGDFAYAAKHAHYHMG 265
QY 242 GIECKFLLEDOPNPVIVQWIDKFWRGYRLTLVAMMDYMQPKRY-----MS 292
DB 266 YLPVVRARGENEPGGVPF-----GYLADICQSSRVNIYEDFVRSLDVVATGAML 314
QY 293 WRESWEMVAEONGALFKDLA-----RYGIREPKWQDACEGK 330
DB 315 YDQIW-LGSYMSGGVGFQYQYATAATDNLLDFTYFGEYVEDKYGCEAPNMTDV--- 370
QY 331 DHISHQAWSTYFGENAASAFHTWPTDEMG-----WLSAKYPSDFR 373
DB 371 --LDVASEVTFYGLQEYEPALL--EDQFGSGQRAAVVAAAAGCSTAFATANAQTGLSG 426
QY 374 YRPRFDHWGQARAGNRFYKTLPLMLCQTQIPMLFTEPGNPTKI-GARSNYLGKFKH 432
DB 427 WYLSMYLHKEQHSRLG--FYGYDLDQDQCGASNVFSIRGDEGLPLEIRGPNPNYAMNVGH 484
QY 433 FCSDHCKDIEDHEPKYVQAWLPVHQIHQGNCFPPDADPAGGDFPLAALVDIYAVTMTGR 492
DB 485 -----QGEYAGISQA-----PHAARGDAFVFNPLV-----KIAFAD 515
QY 493 DNLDFDGSSEDKNEA 507
DB 516 DNLVDFDTNVRGEFA 530
RESULT 12
S74487
hypothetical protein sl11060 - Synecocystis sp. (strain PCC 6803;
C;Species: Synecocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S74487
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74487
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1032 <KAN>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:gl651650; PIDN:BAAL6639.1; PID:d101
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG
Query Match 3.5%; Score 99.5; DB 2; Length 1032;
Best Local Similarity 20.5%; Pred. No. 5.3; Mismatches 50; Indels 75; Gaps 15;
Matches 58; Conservative 50;
QY 54 DPFRLTMDAYKYGKEKKLYAVIDAFTONNAF-----LGVSDA-----RYI-NALKL 101
DB 712 DPLYLWTA---KVGGEHSTLYWLIDAYTTSNYSYSPDGEDANQPGNRNRYINSVKI 767
QY 102 FLQGVY-----PLEYLAHRGFAHVGHRHFTGEGARIACQMSIDELRHYOTE-- 147

Db 768 LVDAYNGDVREFFTDKODPLINAWQKTFPELFPSMPSILKSHIRYPVDMFSTQSERL 827
QY 148 -THAMSTYNKFFNGFHSNQWDFRWWVLSVPKSFEDAYSS-GPFEF--LTAVSFSEFY 202
Db 828 LTYHMEDIDVFN--REDQW-----RPOEYADEQOPIAPYLLIMKLADAKEE 877
QY 203 VLNLFLVPMGSAAYNGDMSTVTFGSAQSDSRH-----MTLGIEICIKELL 250
Db 878 FVLSQVTP-----NARNNLIALLE-ARCEQNYGKLLLYLTPKERVYGPQIEALV 929
QY 251 EODPDNVPVQVWIDKWFWRGYR-----LLTLVAMMDYMQP 287
Db 930 NQD-----PVISERLSLNNRRGSAIQGNLLVPIEESLLYVEP 968
RESULT 13
E28544
Methyl coenzyme M reductase (EC 1.8.-.-) I alpha chain - Methanobacterium thermoautotroph
A:Species: Methanobacterium thermoautotrophicum
A:Variety: strain Marburg
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999
C:Accession: E28544; S78576
R:Bokranz, M.; Baeumner, G.; Allmansberger, R.; Aukel-Fuchs, D.; Klein, A.
J. Bacteriol. 170, 568-577, 1988
A:Title: Cloning and characterization of the methyl coenzyme M reductase genes from Meth
A:Reference number: A91891; MUID:88115150
A:Accession: E28544
A:Molecule type: DNA
A:Residues: 1-550 <BOK>
A:Cross-references: GB:X07794; GB:M18969; NTD:944607; PIDN:CAA30639.1; PID:g44612
A:Experimental source: strain Marburg
R:Rospert, S.; Linder, D.; Ellermann, J.; Thauer, R.K.
Eur. J. Biochem. 194, 871-877, 1990
A:Title: Two genetically distinct methyl-coenzyme M reductases in Methanobacterium therm
A:Reference number: S13864; MUID:91099370
A:Accession: S78576
A:Molecule type: protein
A:Residues: 2-19 <ROS>
A:Experimental source: strain Marburg
R:Ermler, U.; Grabarse, W.; Shima, S.; Goubeaud, M.; Thauer, R.K.
Science 278, 1457-1462, 1997
A:Title: Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological
A:Reference number: A58866; MUID:98035783
A:Contents: annotation; X-ray crystallography, 1.45 angstroms
A:Note: the modified residue designated as N1-methylhistidine is correctly named 3'-meth
C:Complex: heterohexamer of two alpha, two beta (see PIR:A28544) and two gamma (see PIR:
C:Function:
A:Description: catalyzes the reaction of methyl coenzyme M (2-(methylthio)enthanesulfoni
de and methane
A:Pathway: methanogenesis
C:Superfamily: methyl coenzyme M reductase alpha chain
C:Keywords: heterohexamer; metalloprotein; methanogenesis; methylated amino acid; nickel
F:2-550/Product: methyl coenzyme M reductase I alpha chain #status experimental <NAT>
F:447/Binding site: coenzyme F430 nickel (Gln) (axial ligand) #status experimental
F:257/Modified site: 3'-methylhistidine (His) #status experimental
F:271/Modified site: 5-methylarginine (Arg) #status experimental
F:333/Active site: Tyr #status predicted
F:400/Modified site: 2-methylglutamine (Gln) #status experimental
F:444/Binding site: coenzyme M (Tyr) #status experimental
F:445/Modified site: 1-thioglycine (Gly) #status experimental
F:452/Binding site: methyl (Cys) (covalent) #status experimental
F:481/Binding site: coenzyme B (Asn) #status experimental
Query Match 3.4%; Score 99; DB 1; Length 550;
Best Local Similarity 19.5%; Pred. No. 2.6;
Matches 84; Conservative 56; Mismatches 182; Indels 108; Gaps 18;
QY 141 LRHYQTETH---AMSTYKFPNGFHSNQWDFRWWVLSVPKSFED-----AYSSGPF 190
Db 146 VQEHMVETHPALVADSVKVTGNDETADEIDPAFVIDINKQFPEDQAEILKAEVGGIIV 205
QY 191 EFL---TAVSFSEFYVLT---NLLFVPMGSAAYN---GDMSTVTFGSAQSDSRHMTL 241

Db 206 QVVRIPITVSTCDGATTSRWSAMQIGMSMISAYKQAAGEAAIGDFAYAAKHAIEVHMGT 265
QY 242 GIECIKELLEODPDNVPVQVWIDKWFWRGYRLLTLVAMMDYMQPKRV-----MS 292
Db 266 YLPVRRARGENEPGGVPF-----GYLADICQSSRVNYEDPVKPSLDVVAIGAML 314
QY 293 WRESWENYAONGGA-----LFKDLARYG---IREPKGWQDACEGKDHISHQ 336
Db 315 YDQIW-LGYSMSGVGFTQYATAATYDNLDDFTYFGKEYVEDKYGLCEAPNNMDTVLDV 373
QY 337 AWS-TFYGFNAASAFHWVPTEDMG-----WLSAKYPSDFORYRPR 378
Db 374 ATEVTFYGLEQYEYEPALL--EDQFGGSQRAAVVAAAAGCSTAFATGNAQTGLSGWLSM 431
QY 379 FDHWGEQARAGNRFYMKTLPLMTCQIPMLFTEPGNCTKI-GAESNYLKNKHFCSGDH 437
Db 432 YLHKEQHSRLG--FYGYDLQOCGASNVFSTRGDEGLLELRGNPNYAMNVGH----- 484
QY 438 CKDIFDHPQKYVQAWLFPVHQIHQGNCFPPDPADPGAEGFDPLAAVLDYVAVTMGRDNLDF 497
Db 485 -----QGEYAGISQA-----PHAAGDAFVENPLV-----KIAFADDNLVLF 520
QY 498 DGSSEQKNFA 507
Db 521 DFTNVRGEFA 530
RESULT 14
H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: H72228
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: H72228
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-785 <ARN>
A:Cross-references: GB:AF001806; GB:AE000512; NID:g4982196; PID:g4982197; TIGR:TM1624
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1624
Query Match 3.4%; Score 97; DB 2; Length 785;
Best Local Similarity 17.8%; Pred. No. 6;
Matches 86; Conservative 68; Mismatches 147; Indels 190; Gaps 28;
QY 54 DPFRLTMDAYWKY-----QGEKEKKIYAVIDAFTQ---NNAFLGVSDARYI--- 96
Db 49 DLFEIEDREWIYEREFEEKEDKGERVDLVFEGVDLSDVYLVNGVYLGSTEDMFIEYR 108
QY 97 -----NALKLIQ-----GV-----TPLEYLAHRGFAHVGRHFTG-- 126
Db 109 FQVTNVLKEKNHLKVIKSPIRVPKTLQNGYGLGGPDPI-----RGYIRKAQYISYGD 163
QY 127 EGARIACQ-----MQSIDLRHYQTETHAMSTYKFPNGFHH----- 163
Db 164 WGARIVTSGIWKVPYLEVYRARLQDSTAYLLELFGKDALVR-VNGFVHGEGLIVEVYVN 222
QY 164 -----SNQWFDRWYLSVPKSFEDAYSSGPFELTAVSFSEFYVLTNLFV 210
Db 223 GEKIGEPVLEKNEKELFDGVFHLKDVKLWY-----PWNVGPYLYDFVFLKDL--- 272
QY 211 PFMSGAAYNGDMSTVTFGSAQSDSRHMTLGIEICIKELLEQDFDNVP-IQVRWIDKWF 269
Db 273 -----NGEI-----YREKK---ICLRRVRIVQEPDEGKTFIFEINGEKVFA 312

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:28:54 ; Search time 92.32 Seconds
(without alignments)
387.526 Million cell updates/sec

Title: US-09-430-029-5
 Perfect score: 2873
 Sequence: 1 MDTPUUKKGLKDRYAAMT.....EDGSEDOKNFAANRGOATRN 516

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

```
Database :
SPTREMBL_12:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2372	82.6	536	2	Q92NP4	Q92NP4	comamonas t
2	2310.5	80.4	513	2	O52572	O52572	pseudomonas
3	2027	70.6	504	2	O84961	O84961	raistonia s
4	2021	70.3	372	2	Q92NE3	Q92NE3	burkholderi
5	1854.5	64.5	516	2	O52173	O52173	pseudomonas
6	1849.5	64.4	516	2	O52164	O52164	pseudomonas
7	1799	62.6	511	2	O32431	O32431	acinetobact
8	1795	62.5	511	2	O43981	O43981	acinetobact
9	1542	53.7	377	2	Q92454	Q92454	pseudomonas
10	1502	52.3	381	2	Q92455	Q92455	pseudomonas
11	1493	52.0	381	2	Q92457	Q92457	pseudomonas
12	1354	47.1	327	2	Q92456	Q92456	pseudomonas
13	860	29.9	205	2	Q92458	Q92458	pseudomonas
14	844	29.4	204	2	O92NE2	O92NE2	comamonas s
15	795	27.7	215	2	Q92459	Q92459	pseudomonas
16	696.5	24.2	169	2	O30590	O30590	burkholderi
17	386	13.4	500	2	P95408	P95408	pseudomonas
18	377.5	13.1	498	2	O87798	O87798	pseudomonas
19	370	12.9	501	2	O53027	O53027	nocardiac co

20	368.5	12.8	497	2	087082	087082 xanthobacte
21	354.5	12.3	501	2	069178	069178 alcaligenes
22	333.5	11.6	501	2	Q51939	Q51939 burkholderi
23	324	11.3	501	2	007068	007068 burkholderi
24	322	11.2	69	2	Q92NL2	Q92NL2 unidentified
25	320	11.1	69	2	Q92NL4	Q92NL4 unidentified
26	317	11.0	69	2	Q92NL5	Q92NL5 unidentified
27	279	9.7	69	2	Q92NL7	Q92NL7 unidentified
28	279	9.7	69	2	Q92NL1	Q92NL1 unidentified
29	278	9.7	69	2	Q92A36	Q92A36 unidentified
30	271	9.4	69	2	Q92NL6	Q92NL6 unidentified
31	271	9.4	69	2	Q92NL3	Q92NL3 unidentified
32	263	9.2	69	2	Q92NL8	Q92NL8 unidentified
33	227.5	7.9	526	2	Q06116	Q06116 methylocyst
34	162	5.6	233	2	Q944R3	Q944R3 unidentified
35	124.5	4.3	174	2	Q34217	Q34217 uncultured
36	116	4.0	341	2	Q92ET3	Q92ET3 xanthobacte
37	115	4.0	343	2	Q53025	Q53025 nocardia co
38	111.5	3.9	851	12	Q85491	Q85491 bovine leuk
39	111.5	3.9	1416	12	Q92812	Q92812 bovine leuk
40	110	3.8	1017	5	Q9XYV2	Q9XYV2 lymantria d
41	109	3.8	150	2	Q9XB08	Q9XB08 uncultured
42	109	3.8	329	2	Q51943	Q51943 burkholderi
43	107.5	3.7	174	2	Q34218	Q34218 uncultured
44	106.5	3.7	1780	10	Q92T82	Q92T82 arabidopsis
45	99.5	3.5	1742	5	Q61708	Q61708 caenorhabdi

ALIGNMENTS

```

RESULT 1
Q9ZNP4 PRELIMINARY; PRT; 536 AA.
ID Q9ZNP4;
AC 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN APHN.
OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
[1]
RC SEQUENCE FROM N.A.
RP STRAIN=TA441;
RX MEDLINE; 99018839.
RA ARAI H.; AKAHIRA S.; OHISHI T.; MAEDA M.; KUDO T.;
RT "Adaptation of Comamonas testosteroni TA441 to utilize phenol:
RT organization and regulation of the genes involved in phenol
RT degradation.";
RL Microbiology 144:2895-2903(1998).
DR EMBL; AB006479; BAA34172.1; -
SQ SEQUENCE 536 AA; 61860 MW; 29DB0C60 CRC32;

Query Match 82.6%; Score 2372; DB 2: Length 536;
Best Local Similarity 81.4%; Pred. No. 6.8e-197;
Matches 416; Conservative 45; Mismatches 50; Indels 0; Gaps 0

QY 1 MDPTLKKKLGKDRYAAWTRGIGWETTYQPMDKVPYDVEGIKIHWDKWKVDPFLRTM 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MDAPIYKKKLGKLDRTAMTRGLGWETSTOYPMDKVPYDVEGIKIHWDKWKVDPFLRTM 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 DAYWKYQGEKKLYAVIDAFTQNNAFGLGVSDARYINALKFLQGVTFPLEYIAHRGFAHV 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 DAYWKYQGEKKLYAVIEAFAGNNGQLGVTDAARYNALKFLIQGVVPLEYIAHRGFAHA 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 GRFTGEGARIAOMOSIDELRHQYQTHAMSTYKNKFNFGHSHNQWFDRVWYLSVKPSF 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GRFTGAGAKAAQOMOSIDELRHQYQTHALESNYKNKFNFGHSHSHWFDRVWYLSVKPSF 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 181 FEDAYSSGPFPEFLTAVSFSFEYVLNLLFVPPMSGAAYNGDMSTVTFGSAQSDSRHMT 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


DE	PHENOLHYDROXYLASE COMPONENT.
OS	Acinetobacter calcoaceticus.
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC	Moraxellaceae; Acinetobacter.
KN	[1]
RP	SEQUENCE FROM N. A.
EC	STRAIN=NCIB8250;
FX	MEDLINE; 96154937.
RA	EHT S., SCHIRMER F., HILLEN W.;
RT	"Genetic organization, nucleotide sequence and regulation of
RT	expression of genes encoding phenol hydroxylase and catechol 1,2-
RT	dioxygenase in Acinetobacter calcoaceticus NCIB8250."
RL	Mol. Microbiol. 18:13-20(1995).
DR	EWBL; Z36903; CAA85383.1; -
SQ	SEQUENCE 511 AA; 60316 MW; DA052BCE CRC32.

Query Match 62.5%; Score 1795; DB 2; Length 511;
Best Local Similarity 64.1%; Pred. NO. 4.5e-147;
Matches 325; Conservative 60; Mismatches 112; Indels 10; Gaps 2;

QY	8	KKGLKDRYAA	MTGLGWETTYQ	PMDKVFFPYDRY	EGIKIHDKWQWD	PERLTMDAYWKYQ	67
Db			:	:	:	:	:
QY	13	KKLNAKERYIL	TRDLWDSEYAD	RDKDAPYEEF	EGIKITDQSKWED	PFLIMDNTWKYQ	72
Db			:	:	:	:	:
QY	68	GEKEKKIYAVI	DAFTONNAFLG	VSDDARYINAKL	FLFGVTPLEYLA	HRGFHAGFVHGRHTE	127
Db			:	:	:	:	:
QY	73	AEKEKKIYAI	FAQNGQMNVS	ERYNAIKLEF	TAVTPLEYQAY	GVHVGROFSGI	132
Db			:	:	:	:	:
QY	128	GARIACOMQSI	DELHRHVOTTH	AMSTYNNKFP	NGFPHSNQWDF	RVWLVSPKSFEDAYS	187
Db			:	:	:	:	:
QY	133	GARIASOMQSI	DELHRHVOTQI	HAMSHYNK	FFDGFQWAMHMD	HRVLSVPKSFEDARSA	192
Db			:	:	:	:	:
QY	188	GPEEFLTAVS	FPEYVLTNLL	FPVPMGSAAY	NGDMSTVTTF	GFSAQSDSRHMTLGL	247
Db			:	:	:	:	:
QY	193	GPEEFLLAIS	FPEYVLTNLL	FPVPMGSAAY	NGDMAVTTF	GFSAQSDSRHMTLGL	252
Db			:	:	:	:	:
QY	248	FILEQPDNDP	PIVORWIDK	WFRGRLTL	VLAMMDMYQF	KRYMSWRESWEYAEONGGA	307
Db			:	:	:	:	:
QY	253	FILEQHEQND	PVIVQEWIDK	WFRGTRLLS	IVGMMDMYLP	KNKYSWKAEWIFFEAGGA	312
Db			:	:	:	:	:
QY	308	LFKDLARYG	IREPKGWQD	CEGKDHLSHQ	AWSTFYCFNA	SAFHWVPTDEMGWISAKY	367
Db			:	:	:	:	:
QY	313	LFKDLRGIR	MPKPKYSE	VTSKEKEHSHQ	AWWLFYNFGH	AAGFHTWIPIDEEMDLSEKY	372
Db			:	:	:	:	:
QY	368	PDSEFDRYR	PRFDHWC	EQARAGNRFY	MKTLPMLCQFCQ	IPMLFETPGNPTKIGARESNY	426
Db			:	:	:	:	:
QY	373	PDFDKYRPR	WELARKM	EAEGKRFY	SAGLUPQCVCQ	IPMTIFTEMDGPTTLFYSYRDSIY	432
Db			:	:	:	:	:
QY	427	LGNKFHCS	DCKDIFD	HEPQKYQAWL	VPVHQIHOCN	CFPPADPAEGAFDPLAAVL	486
Db		:	:	:	:	:	:
QY	433	KDERYHTCS	DGCHDIF	EREPEKYQ	IAWLPVHQI	LQNCGGPDLE-----SIIRDY	483
Db		:	:	:	:	:	:
QY	487	AVTMGRDNL	DFDGS	EQKFNFA	WRGGA	513	
Db		:	:	:	:	:	:
QY	484	NFVGNADNL	IEGSPDQ	QRKWKW	KNNA	510	
Db		:	:	:	:	:	:

RESULT	9
Q92454	
ID	Q92454;
AC	PRELIMINARY; PRT; 377 AA.
DT	01-MAY-1999 (TRENBLUREL. 10, Created)
DT	01-MAY-1999 (TRENBLUREL. 10, Last sequence update)
DT	01-NOV-1999 (TRENBLUREL. 12, Last annotation update)
DE	PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
[1]
RN RN SEQUENCE FROM N.A.
RP STRAIN=P-10;
RC
RA FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;

RT	"Unique primary structure found in phenol hydroxylases Exhibiting high affinity towards trichloroethylene.";
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AB016860; BAA74745.1; -.
FT	NON_TER 1
FT	NON_TER 377
SQ	SEQUENCE 377 AA: 442669 MW: 78668F49 CPO32.

Query Match 53.7%; Score 1542; DB 2; Length 377;
Best Local Similarity 71.6%; Pred. No. 2.1e-125;
Matches 270; Conservative 46; Mismatches 61; Indels 0; Gaps 0;

QY	63	YWKYQGEKEKLYAVIDAFQNNNAFLCVSDARYINALKFLQGVTFLEYLAHRGFABHVR	122
Db		1 YWKYQSEKERKLYAIDSFYQNNNGHNLNVSDPRYINALRFLTGVTLEYAAHRGYAHLGR	60
QY	123	HFTCEGARIACQMSIDELRHYOTETHAMSTYKNFNGFHHNSQWEDRWVYLSVPKSPFE	182
Db		1 HFTCEGARIACQMSIDELRHYOTETHAMSTYKNFNGFHHNSQWEDRWVYLSVPKSPFE	120
QY	61	HFGCAGARVAACQMSIDELRHAQTQHLTVSYNKNFNGEGEWHMDRWVYLSVPKSYFE	120
Db		1 HFGCAGARVAACQMSIDELRHAQTQHLTVSYNKNFNGEGEWHMDRWVYLSVPKSYFE	120
QY	183	DAYSSGFEFLTAVSFSEFYVLNLLFVPPMSCAAAYNGDMSTVTFGFSQAQSDSRHWTLG	242
Db		1 DAYSSGFEFLTAVSFSEFYVLNLLFVPPMSCAAAYNGDMSTVTFGFSQAQSDSRHWTLG	242
QY	121	DAMSAGFEFITALSFSEFYVLNLLFVPPMSCAAAYNGDMATVTFGFSQAQSDSRHWTLG	180
Db		1 DAMSAGFEFITALSFSEFYVLNLLFVPPMSCAAAYNGDMATVTFGFSQAQSDSRHWTLG	180
QY	243	IECIKFLLEDODPNVPIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWRESWEMYAE	302
Db		1 IECIKFLLEDODPNVPIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWRESWEMYAE	302
QY	181	LEVYKFLCEQDPGNIPLLQKWLKDQWVRGPRLLTVGMMDYMLPKRVMSWAEAWEMYFE	240
Db		1 LEVYKFLCEQDPGNIPLLQKWLKDQWVRGPRLLTVGMMDYMLPKRVMSWAEAWEMYFE	240
QY	303	QNGGALFKDLARYGIRPEPKQWQDACFGKDHISHQAQSTVTFYFNAAAFHTWPTDEMGW	362
Db		1 QNGGALFKDLARYGIRPEPKQWQDACFGKDHISHQAQSTVTFYFNAAAFHTWPTDEMGW	362
QY	241	QAGGALFKDLERYGLRMPKYHEVATYTKDRITHEAGTGYNTAAAAAGFTWVPKQDEMAW	300
Db		1 QAGGALFKDLERYGLRMPKYHEVATYTKDRITHEAGTGYNTAAAAAGFTWVPKQDEMAW	300
QY	363	LSAKYPDSDFRYYRPRFDHWHGQARAGNRFYMKLTPLMTCQTQIPLMFLTEPGNPTKIGAR	422
Db		1 LSAKYPDSDFRYYRPRFDHWHGQARAGNRFYMKLTPLMTCQTQIPLMFLTEPGNPTKIGAR	422
QY	301	LACKYPTETERYKYPLRDHWHQERQOQGEGERFYNATLPLMTCQTQIPLMFLTEPGNPTKIGAR	360
Db		1 LACKYPTETERYKYPLRDHWHQERQOQGEGERFYNATLPLMTCQTQIPLMFLTEPGNPTKIGAR	360
QY	423	ESNYLGNKHFHFCSDHCK 439	
Db		1 ESSYHGMMKFHFCSDGCK 377	

RESULT	10
09Z455	

ID	C92455	PRELIMINARY;	PRT;	381 AA.
AC	C92455;			
DT	01-MAY-1999	(TREMBLrel. 10, Created)		
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).			
OS	<i>Pseudomonas putida</i> .			
OC	Bacteria; Proteobacteria; gamma subdivision; <i>Pseudomonas</i> group;			
OC	<i>Pseudomonas</i> .			

RP
SEQUENCE FROM N. A.

RC STRAIN=P-8;

RA FUTAMATA H.

RT "Unique pri

RT affinity to

RL Submitted (

DR EMBL; AB016

NON_TER

NON_TER

SQ SEQUENCE

2
 3
 4
 5
 6
 7
 8

Query Match

BEST LOCAL SIM
 Metaphor 200

1807
MILITARY

[illegible]

QY 01 DAYWY

1 0701344

TYM T 22

```

QY 121 GRHFTGEGARIACOMSIDELRHVQYOTETHAMSTYNKFNFGFHSNQWEDRWYLSVPKSF 180
DB 61 GROFSGAGARVACQMAIDELRHVQYOTETHAMSHYNKFNHGLHDFAHMHDWRWFLSVPKSF 120
QY 181 FEDAYSSGPFELTAVSFSEFVLTNLLFVPMGAAAYNGDMSTVTFGFSQSDSRHMT 240
DB 121 FEDARTAGPFELTAVSFSEFVLTNLLFVPMGAAAYNGDMATVTFGFSQSDSRHMT 180
QY 241 LGIECIKFLLEQDPDNPVIVQWIDKWFWRGYRLTLVAMMDYMQPKRVMWSRESWEMY 300
DB 181 LGLEVIKFLLEQHDNPVIVQWIDKWFWRGYRLTLVAMMDYMLPNKVMWSRESWEMY 240
QY 301 AEONGCALFKDLARYGIRPKKVEEANIADKHSQAWSTFYGNAASAFHTWPTDEEM 360
DB 241 FEEAGALFKDLARYGIRPKKVEEANIADKHSQAWSTFYGNAASAFHTWPTDEEM 300
QY 361 GWLSAKYPDSFDRYRPRFDHGWGEQARAGNRFYMKTLPMLCQTCQIPMLFTEPGNPTKIG 420
DB 301 DMLSSKYPDTFKIYRPRYEHNRALQEKGERFYNPTLPMLCQICQIPLFSGEPDDTTLS 360
QY 421 ARESNYLGKHFHCSHCKDI 441
DB 361 HRSVHEGERYHFCSGGCCDI 381

RESULT 11
Q92457
ID Q92457 PRELIMINARY; PRT; 381 AA.
AC Q92457;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P-5;
RA FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;
RT "Unique primary structure found in phenol hydroxylases Exhibiting high
RT affinity towards trichloroethylene."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016857; BAA7473.1; -
DR EMBL; AB016854; BAA7473.1; -
FT NON_TER 1
FT NON_TER 381
SQ SEQUENCE 381 AA; 44525 MW; 5591FE88 CRC32;

Query Match 52.0%; Score 1493; DB 2; Length 381;
Best Local Similarity 70.1%; Pred. No. 3.7e-121;
Matches 267; Conservative 43; Mismatches 71; Indels 0; Gaps 0;

QY 61 DAYWKYGEKKIYAVIDAFTONNAPLGVSDARYINALKFLQGVTPLEYLAHRGFAHV 120
DB 1 DTWKYQAEKKIYAFIDAFSONNGHTLSDARYNALKFLSGVTPLEYQAYQGFARV 60
QY 121 GRHFTGEGARIACOMSIDELRHVQYOTETHAMSTYNKFNFGFHSNQWEDRWYLSVPKSF 180
DB 61 GROFSGAGARVACQMAIDELRHVQYOTETHAMSHYNKFNHGLHDFAHMHDWRWFLSVPKSF 120
QY 181 FEDAYSSGPFELTAVSFSEFVLTNLLFVPMGAAAYNGDMSTVTFGFSQSDSRHMT 240
DB 121 FEDARTAGPFELTAVSFSEFVLTNLLFVPMGAAAYNGDMATVTFGFSQSDSRHMT 180
QY 241 LGIECIKFLLEQDPDNPVIVQWIDKWFWRGYRLTLVAMMDYMQPKRVMWSRESWEMY 300
DB 181 LGLEVIKFLLEQHDNPVIVQWIDKWFWRGYRLTLVAMMDYMLPNKVMWSRESWEMY 240
QY 301 AEONGCALFKDLARYGIRPKKVEEANIADKHSQAWSTFYGNAASAFHTWPTDEEM 360
DB 241 FEEAGALFKDLARYGIRPKKVEEANIADKHSQAWSTFYGNAASAFHTWPTDEEM 300

RESULT 13
Q92458
ID Q92458 PRELIMINARY; PRT; 205 AA.
AC Q92458;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).
OS Pseudomonas putida.

```

```

DB 241 FEEAGALFKDLARYGIRPKKVEEANIADKHSQAWSTFYGNAASAFHTWPTDEEM 300
QY 361 GWLSAKYPDSFDRYRPRFDHGWGEQARAGNRFYMKTLPMLCQTCQIPMLFTEPGNPTKIG 420
DB 301 DMLSSKYPDTFKIYRPRYEHNRALQEKGERFYNPTLPMLCQICQIPLFSGEPDDTTLS 360
QY 421 ARESNYLGKHFHCSHCKDI 441
DB 361 HRSVHEGERYHFCSGGCCDI 381

RESULT 12
Q92456
ID Q92456 PRELIMINARY; PRT; 327 AA.
AC Q92456;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P-6;
RA FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;
RT "Unique primary structure found in phenol hydroxylases Exhibiting high
RT affinity towards trichloroethylene."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016858; BAA7474.1; -
FT NON_TER 1
FT NON_TER 327
SQ SEQUENCE 327 AA; 38388 MW; DFF0A7B0 CRC32;

Query Match 47.1%; Score 1354; DB 2; Length 327;
Best Local Similarity 74.6%; Pred. No. 3e-109;
Matches 244; Conservative 31; Mismatches 52; Indels 0; Gaps 0;

QY 65 KYQGEKKIYAVIDAFTONNAPLGVSDARYINALKFLQGVTPLEYLAHRGFAHVGRHF 124
DB 1 KYQAEKKIYAFIDAFQNNQNGHNTSDARYNALKFLSGVTPLEYQAYQGFARVGRHF 60
QY 125 TGGARIACOMSIDELRHVQYOTETHAMSTYNKFNFGFHSNQWEDRWYLSVPKSFEDA 184
DB 61 GGAGARVACQMAIDELRHVQYOTETHAMSHYNKFNHGLHDFAHMHDWRWFLSVPKSFEDA 120
QY 185 YSSGPFELTAVSFSEFVLTNLLFVPMGAAAYNGDMSTVTFGFSQSDSRHMTLGIE 244
DB 121 RTAGPFELTAVSFSEFVLTNLLFVPMGAAAYNGDMATVTFGFSQSDSRHMTLGIE 180
QY 245 CIKFLLEQDPDNPVIVQWIDKWFWRGYRLTLVAMMDYMQPKRVMWSRESWEMYAFQN 304
DB 181 VIKFLLEQHDNPVIVQWIDKWFWRGYRLTLVAMMDYMLPNKVMWSRESWEMYFEQA 240
QY 305 GGALFKDLARYGIRPKKVEEANIADKHSQAWSTFYGNAASAFHTWPTDEMGWLS 364
DB 241 GGALFKDLARYGIRPKKVEEANIADKHSQAWSTFYGNAASAFHTWPTDEMGWLS 300
QY 365 AKYPDSDRYPREDHGWGEQARAGNR 391
DB 301 AKYPDSDRYPREDHGWGEQARAGNR 327

RESULT 13
Q92458
ID Q92458 PRELIMINARY; PRT; 205 AA.
AC Q92458;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).
OS Pseudomonas putida.

```


OC Bacteria; Proteobacteria; gamma subdivision; pseudomonas group;
OC Pseudomonas.

RN	[1]	
RP	SEQUENCE FROM N. A.	
RC	STRAIN=P-4;	
RA	FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;	
RT	"Unique primary structure found in phenol hydroxylases Exhibiting high	
RT	affinity towards trichloroethylene.";	
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AB016856; BAA74741.1; -.	
FT	NON_TER	1
FT	NON_TER	205
FT	SEQUENCE	205
SQ	205 AA; 23441 MW; EI9D7C89 CRC32;	

Query Match 29.9%; Score 860; DB 2; Length 205;
Best Local Similarity 78.5%; Pred. No. 7.6e-67;
Matches 161; Conservative 17; Mismatches 27; Indels 0; Gaps 0

```

55  PFRUTDAYWKYQEKEKKLYAVIDAFTQNNAFLGVSDARYNALKFLQGVTTPLEYLAH 114
Db  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1  PFRUTMDTYWKYQAEKEKKLYAIDAFSONNGHTTSLDARYNALKFLSGVTTPLEYQAY 60
QY  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 RGFHVHGHTGEGARIACQMSIDELRHYOTETHAMSTYKFFNFNGFHSNOWNDFRWYL 174
Db  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61  QGFARVGQFSGGARIACQCAIDELRHVQTIHAMSHYKHNGLHDFAHMHDRVWFL 120
QY  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 SVPKSFEDAYSSGPFPEFLTAVSFEVYLINLLFVPPMSGAAYNGDMSTVTFGSAQSD 234
Db  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 SVPKSFEDARTAGPFPEFLTAISFEVYLINLLFVPPMSGAANFNGDMATVTFGSAQSD 180

```

QY 235 ESRHMTLGIECIKELLEQDPDNVPI 259
| : | | | | | : | | | | | | | | |
Db 181 EARHMTLGLGVIKFLEEQHEDNVPI 205

RESULT	14
Q9ZNE2	
ID	Q9ZNE2 PRELIMINARY; PRT: 204 AA.
AC	Q9ZNE2;
DT	01-MAY-1999 (TREMBlrel. 10, Created)
DT	01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE	PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).
OS	Comamonas sp.
OC	Comamonas
EC	Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-E6;
RA	FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;
RT	"Unique primary structure found in phenol hydroxylases Exhibiting high affinity towards trichloroethylene.";
RT	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL	EMBL; AB016862; BAA74747.1; -.
FT	NON_TER 1
DR	NON_TER 1
FE	NON_TER 204
SQ	SEQUENCE 204 AA; 23298 MW; DB5A9322 CRC32;

Query Match 29.4%; Score 844; DB 2; Length 204;
Best Local Similarity 74.0%; Pred. No. 1.8e-65;
Matches 151; Conservative 27; Mismatches 26; Indels 0; Gaps 0;

QY	57	R L M D A Y T W K Q G E K K L Y A V I D A F T O N P A F L G V S D A R Y I N A L K F L Q G V P L E Y L A H R G	116
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
D b	1	R M T R D S Y K Y C A E K R K L Y A I I D A F T O N N G L H G V T D A R Y I N A L K F L T G V S P L E Y M A H R G	60
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	117	F A H W G R U F T C E G A R I A C O M O S I D E L R H V O T H A M S T Y N K F E F H S N Q W F D R V W Y L S V	176
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
D b	61	F A H W G R F P G A G P V A C L M O S L D E I R A Q T I H S L S Y N K H Y N G F A N W R H Q H D R V W Y L S V	120
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	177	P K S F F E D A Y S G G P E F F T A V S F S F E Y V L T L L F P F M S G A A Y N G D M S T V T F G F S A Q S D E S	236
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	

```

Db      121  PKSFDDAFTAGPFEFMIAIGFAFAYVLTLLFVPFISGAAYNGDMGAMAFGFSAGSDSES 180
Qy      237  RHMTLGIECIKFLLEQDPDNVPV 260
          |||||:||||:|||||:||||
Db      181  RHMTLGLEIKFLLEQDPDNLPV 204

RESULT 15
Q9Z459
ID      Q9Z459          PRELIMINARY;      PRT;      215 AA.
AC      Q9Z459;
DT      01-MAY-1999 (T-EMBLrel. 10, Created)
DT      01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT      01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE      PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).
OS      Pseudomonas putida.
OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC      Pseudomonas.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=P-2;
RA      FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;
RT      "Unique primary structure found in phenol hydroxylases Exhibiting high
RT      affinity towards trichloroethylene."
RL      Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AB016855; BAA74740.1; -.
FT      NON_TER      1
FT      TER      215      215
SQ      SEQUENCE      215 AA; 25713 MW; AEBD075A CRC32;

```

Query Match	27.7%;	Score 795;	DB 2;	Length 215;
Best Local Similarity	64.0%;	Pred. No. 3.3e-61;		
Matches 137;	Conservative 29;	Mismatches 48;	Indels 0;	Gaps 0;
QY	233	SDSRHMTLGLCTKFFLEQDDNPVIVORWIDKFWNGYRLLITVAMMMDYMPKVMKMS	292	
Db	1	SDSRHMTLGLGVTKFFLEQEDNPVIVQRIQWIDKFWNGYRLLITVGMMDYMLPNKVMKMS	60	
QY	293	WRESWEYAFONGALPKDILARYGIREPKGMQDACECKDHISHQAWSTFYGFNAASAFHT	352	
Db	61	WSEAWGYFEEAGALPKDLERYGRTPPKHVEEANIAKDHVSHQAWSIFYQYSGATNFT	120	
QY	353	WVPTDEDMGLWSAKYPPSDFRYIYPRPDHWCEQARAGNRYMKTPLMLCQCIPMLCFTE	412	
Db	121	WMPDDELDWLSEKYPPTFDKIYPRPEHFWALQEKERYFNPTPLMLCQCIPQLSFG	180	
QY	413	PGNPFTKIGARESNYLGNKHFHFCSDHCKDFIDFHEP	446	
Db	181	PDDQTTLTSHRSVHGEGRYHFCSGCCDIFPEY	214	

Search completed: September 26, 2000, 20:28:57
Job time: 1016 sec

Db 181 FEDATGPPFEFLTAVSFSEYVLTNLLFVPMFGAAYNGDMSTVTFGSAQSDSRHMT 240
 QY 241 LGTECKFLEQPDNPVIVQWIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEY 300
 Db 241 LGTECKFLEQPDNPVIVQWIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEY 300
 QY 301 AEONGALPKOLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTVWPTDEDM 360
 Db 301 VEONGALKERDLARYGIRPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTVWPTDEDM 360
 QY 361 GWLSAKYPSDFRYPREDHWGEQARAGNRFFMKTLPMLCOTQCPMLFTEPGNFKIG 420
 Db 361 GWLSAKYPSDFRYPREDHWGEQARAGNRFFMKTLPMLCOTQCPMLFTEPGNFKIG 420
 QY 421 ARESNVLGNKHFCSCHDKDIEDHEPKYQVQAWLPVHQIHQGNCFPPDADPGAEGEDPLA 480
 Db 421 YRESAYLGDYHFCSDHCKEIDFNEPEKEVQSWLPVQVYQGNCFPPDADPGAEGEDPLA 480
 QY 481 AVLDYAVTMGRNLDLFDGSEQKNAWRG 512
 Db 481 ALLDIYNLVGRNDFDEGSEQKNAWRG 512

RESULT 2

W98973
 ID W98973 standard; Protein: 504 AA.
 AC W98973;
 DT 10-MAY-1999 (first entry)
 DE Alcaligenes sp. protein PoxD.
 KW Alcaligenes; PoxR; PoxA; PoxB; PoxC; PoxD; PoxE; PoxF; PoxG; PoxH;
 KW PoxI; aromatic; oxidative; petroleum purification; chemical industry;
 KW drug.
 OS Alcaligenes sp.
 PN J11042088-A.
 PD 16-FEB-1999.
 PF 25-JUL-1997; 200625.
 PR 25-JUL-1997; JP-200625.
 PA (TOFU) TONEN CORP.
 DR WPI: 99-197820/17.
 DR N-PSDB: X18867.
 PT New encoding an aromatic cpd. oxidative decompsn. enzyme - useful in
 the fields of petroleum purification, chemical and drug industries
 PS Claim 1: Page 7-19; 35pp; Japanese.
 CC The present sequence encodes Alcaligenes sp. protein PoxD. The present
 invention also describes PoxA, PoxB, PoxC, PoxE, PoxF, PoxG, PoxH
 and PoxI. The proteins are useful in the fields of petroleum
 purification, chemical industry and drug industry related to the
 synthesis, conversion and decomposition of aromatic compounds.
 CC Sequence 504 AA;
 SQ

Query Match 70.6%; Score 2027; DB 1; Length 504;
 Best Local Similarity 70.3%; Pred. No. 2e-201;
 Matches 355; Conservative 62; Mismatches 78; Indels 10; Gaps 1;
 QY 7 KKLGLKDRYAAWRGLGWETTYQPMOKVFPYDREYEGIKHDKWDVDFRLTMDAYWKY 66
 Db 4 RKKLNLREKATMDRGLGWETTYQPMOKVFPYDREYEGIKHDKWDVDFRLTMDAYWKY 63
 QY 67 QGEKELIAYIDAFTQNNAPFLGVSDARYINALKFLQGVTPLEYLAHRGFVHGRHFTG 126
 Db 64 QSEKERLIAYIDSFVQNGHNLVSDARYINALKFLQGVTPLEYLAHRGFVHGRHFTG 123
 QY 127 EGARIACOMSIDELRHQYQETHAMSTYKFFNGFHHSNQWDFRVVYLSVPKSFEDAYS 186
 Db 124 AGARVAAQMSIDELRHAQQLHLSYVNYKFFHGFGEGRHMDRVVYLSVPKSFEDAYS 183
 QY 187 SGPEFLTAVSFSEYVLTNLLFVPMFGAAYNGDMSTVTFGSAQSDSRHMTLGIECI 246
 Db 184 AGFEFLTAVSFSEYVLTNLLFVPMFGAAYNGDMSTVTFGSAQSDSRHMTLGIECI 243
 QY 247 KFLLEQPDNPVIVQWIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEYAEONGG 306
 Db 247 KFLLEQPDNPVIVQWIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEYAEONGG 306

Db 244 KFLCEQPDNPILPLQKWLDKWFRGRLITLVGMMDYMLPRRYMSWAEAEWTFEQAGG 303
 QY 307 ALPKDLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTVWPTDEMGWLSAK 366
 Db 304 ALPKDLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTVWPTDEMGWLSAK 363
 QY 367 YPDSFDRYPRPDHWGEQARAGNRFFMKTLPMLCOTQCPMLFTEPGNFKIGARESNY 426
 Db 364 YPOTFARYPRPDHWGEQARAGNRFFMKTLPMLCOTQCPMLFTEPGNFKIGARESNY 423
 QY 427 LGNKFHFCSDHCKDIEDHEPKYQVQAWLPVHQIHQGNCFPPDADPGAEGEDPLAALVDY 486
 Db 424 HGMKFHFCSDGCKDIEDHEPKYQVQAWLPVHQIYQGNCFPPDADPGAEGEDPLAALVDY 486
 QY 487 AVTMGRNLDLFDGSEQKNAWRG 511
 Db 474 RLNLGADNLDFEGSDQKNNAWKG 498

RESULT 3

R66215
 ID R66215 standard; Protein: 501 AA.
 AC R66215;
 DT 04-AUG-1995 (first entry)
 DE Nocardia corallina alkene monooxygenase subunit-2.
 KW alkene monooxygenase; subunit-2; epoxidation; amoC.
 OS Nocardia corallina.
 PN J06292571-A.
 PD 21-OCT-1994.
 PF 06-APR-1993; 105171.
 PR 06-APR-1993; JP-105171.
 PA (NIHA) JAPAN ENERGY CORP.
 DR WPI: 95-009069/02.
 DR N-PSDB: Q79569.
 PT Alkene monooxygenase and corresp. gene - useful for the
 epoxidation of an alkene
 PS Claim 1: Page 3-4; 30pp; Japanese.
 CC E.coli transformed with the DNA sequence Q79569 are able to
 catalyze the epoxidation of alkenes. The DNA is derived from
 Nocardia corallina and comprises 4 open reading frames. ORFs amoA
 and amoC encode subunits 1 and 2 of the alkene monooxygenase enzyme
 and ORF amoD encodes a reductase capable of transferring electrons
 from NADH coenzyme to a monooxygenase.
 CC Sequence 501 AA;
 SQ

Query Match 12.9%; Score 370; DB 1; Length 501;
 Best Local Similarity 25.6%; Pred. No. 7.9e-30;
 Matches 124; Conservative 68; Mismatches 190; Indels 102; Gaps 21;
 QY 37 PYDRYEGIKIHDWD-----KWDVDFRLTMDAYWKYQGEKELIAYI 78
 Db 5 PTQLHEKSKSYDMDFTSVERRPKFKETKYMPPKGGKDPFVLRIDYMKAEKDRTHGFL 64
 QY 79 DAFTQNNAPFLGVSDARYINALKFLQGVTPLEYLAHRGFVHGRHFTGEGARIACOMSI 138
 Db 65 DGAVTREATRI-EPFRAEMKIMVPLQTNAYEQAVACCGMILISAVENQELRQGYAAQML 123
 QY 139 DELRHQYQETHAMSTYKFFNGFHHSNQWDFRVVYLSVPKSF----- 180
 Db 124 DEVRHAQLE---MTLRN-----YYAKHWCD-----PSGFDIGORGLYOHAPGLVIG 167
 QY 181 -FEDAYSSGPFEFLTAVSFSEYVLTNLLFVPMFGAAYNGDMSTVTFGSAQSDSRHMT 239
 Db 168 EFQIFNTGDPDLDVITIDUNIVAEATFTNLLVATPQVAVANGDNAMASVFLTQSDARHM 227
 QY 240 TLGIECKFLEQPDNPVIVQWIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEY 298
 Db 228 ANGYGSVMALLENE-DNLPLLNQSLDRHFWRAHKAIDNAVGCWCSEYGARKRWSIKAGWE 286
 QY 299 MYAEQN--GGALFKDLARYGIREPKGWQDACEG-----KDHISHQAWSTFYGFNAAS 348
 Db 287 EWVVDVFGVGYIDR-LSEFGVQAP-----ACLGAAADEVKWSHHTLIGVLSAVVPLN--- 337

```

QY 349 AFHTWVPTEDMG-----WLSAKYPDSFDRYRFRFDHW-GEQA---RAGNRYFMYKTLPM 399
Db 338 ---FW---RSDAMGFADFEWFENHYPGSAAYQ-----ACLGAAAEVKNSHHPLGQVLSAVWPLN--- 337
QY 400 L---CQTCQIPMLFTEPCNPTKIGARESNYLGKHFHFCSDHCKDIFDHEPKY---VQAW 453
Db 389 LPPMCQVCQVPCVMP---RLDMNAARILIEFGQIALCSEPCQRLFTNWPEAYRHRQYW 445
QY 454 LPVH 457
Db 446 ARYH 449

RESULT 4
ID R81471 standard; Protein; 501 AA.
AC R81471;
DE 07-AUG-1996 (first entry)
KW Nocardia corallina alkene mono-oxygenase gene product, amoC.
KW Alkene mono-oxygenase; indole; indigo production; biosynthesis;
OS Nocardia corallina B-276.
FH Key Location/Qualifiers
FT misc_difference 268
FT /note= "Tip residue given in the specification
FT corresponding to a TCG codon in T17418"
PN J08023988-A.
PD 30-JAN-1996.
PF 08-JUL-1994; 179688.
PR 08-JUL-1994; JP-179688.
PA (NIHA ) JAPAN ENERGY CORP.
DR WPI: 96-133426/14.
DR N-PSDB: T17418.
PT Prep'n. of indigo by a microbiological method - by culturing a
PT microbe having alkene monooxygenase activity to oxidise indole to
PT indigo
PS Claim 5; Page 7-8; 11pp; Japanese.
CC R81469-R81472 are protein products of the Nocardia corallina strain
CC B-276 alkene mono-oxygenase gene. The gene encodes 4 protein products
CC amoA, amoB, amoC and amoD derived from the 3 different reading frames
CC of the operon. The gene is useful for the production of indigo via
CC oxidation of indole. Nocardia corallina can be cultured in a medium
CC contg. indole and will readily oxidise the indole yielding indigo into
CC the culture medium. E. coli may also be transformed with the alkene
CC mono-oxygenase gene and used as above to efficiently produce indigo
CC by microbial oxidation.
SQ Sequence 501 AA;

Query Match 12.9%; Score 370; DB 1; Length 501;
Best Local Similarity 25.6%; Pred. No. 7.9e-30;
Matches 124; Conservative 68; Mismatches 190; Indels 102; Gaps 21;

QY 37 PYDRYEGIKTHDWD-----KWVDFRLTMDAYWKYQCEKELKAVI 78
Db 5 PTQLHEKSKSYDWDFTSVERRPKETKYKPKKGKDFRVLIRDMKMEAFKDRTHGFL 64
QY 79 DAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHGRFAHVGHRHFTGEGARIACQMOSI 138
Db 65 DGAVRTREATRI-EPRFAEMKIMVOLTNAEYQAVAGCGMISAVENQELRQGYAAQML 123
QY 139 DELRHYOTETHAMSTYKFFNGPHHSNQWDFRWYLSVPKSF----- 180
Db 124 DEVRHAQLE---MTLRN-----YYAKHWCD-----PSGFDIGQRLYQHAGLYSIG 167
QY 181 -FEDAYSSGPFEEITAVSFSEYVFLNLLFPFMMSGAYNCDMSTVTFGSAQSDSRHM 239
Db 168 EFQHEHTGDLVDLIIDNIVAEATAFTNLLVATPQAVANGDNAMASVFLSIQSDEARHM 227
QY 240 TLGIEICKLLEQDPNPVIVORWIDKWRGYRLT-TIVAMMDYMQPKRVMWSRESWE 298
Db 228 ANGYGSMALLENE-DNLPLLSQSLDRHFWRAHKALDNVAGWCSEYGAARKPWSYKAQWE 286

```

```

QY 299 MYAEQN--GGALFKDLARYGIREPKGWODACEG-----KDHISHQAWSTFFYGFNAAS 348
Db 287 EWVDDDFVGGYIDR-LSEFGVQAP-----ACLGAAAEVKNSHHPLGQVLSAVWPLN--- 337
QY 349 AFHTWVPTEDMG-----WLSAKYPDSFDRYRFRFDHW-GEQA---RAGNRYFMYKTLPM 399
Db 338 ---FW---RSDAMGFADFEWFENHYPGSAAYQ-----ACLGAAAEVKNSHHPLGQVLSAVWPLN--- 337
QY 400 L---CQTCQIPMLFTEPCNPTKIGARESNYLGKHFHFCSDHCKDIFDHEPKY---VQAW 453
Db 389 LPPMCQVCQVPCVMP---RLDMNAARILIEFGQIALCSEPCQRLFTNWPEAYRHRQYW 445
QY 454 LPVH 457
Db 446 ARYH 449

RESULT 5
ID R66213 standard; Protein; 342 AA.
AC R66213;
DE 04-AUG-1995 (first entry)
KW Nocardia corallina alkene mono-oxygenase subunit-1.
KW alkene monooxygenase; subunit-1; epoxidation; amoA gene.
OS Nocardia corallina.
PN J06292571-A.
PD 21-OCT-1994.
PF 06-APR-1993; 105171.
PR 06-APR-1993; JP-105171.
PA (NIHA ) JAPAN ENERGY CORP.
DR WPI: 95-005069/02.
DR N-PSDB: Q79569.
PT Alkene monooxygenase and corresp. gene - useful for the
PT epoxidation of an alkene
PS Claim 1; Page 2; 30pp; Japanese.
CC E.coli transformed with the DNA sequence Q79569 are able to
CC catalyse the epoxidation of alkenes. The DNA is derived from
CC Nocardia corallina and comprises 4 open reading frames, ORFs amoA
CC and amoC encode subunits 1 and 2 of the alkene monooxygenase enzyme
CC and ORF amod encodes a reductase capable of transferring electrons
CC from NADH coenzyme to a monooxygenase.
SQ Sequence 342 AA;

```

```

Query Match 4.0%; Score 115; DB 1; Length 342;
Best Local Similarity 19.8%; Pred. No. 0.0011;
Matches 49; Conservative 39; Mismatches 117; Indels 42; Gaps 6;

QY 44 IKIHWDKWKVDPFRLTMDAYWKYQCEKELKAVIADFTQNNAFVGVSDARYINALKFL 103
Db 58 VRTSWSAYRDPHQLWQRPYVSTCNQDQQLARLVPVLTGMSA----- 110
QY 104 QGVTPLE-----EYLAHGRFAHVGHRHFTGEGARIACQMOSIDELRHYOTE 147
Db 111 -AUIPWISQILARSYAANPVEYGLFSLAYAVQAQMSDVFQFVQVADRMELLDQI 169
QY 148 THAMSTYKNF--FNGFHHSNQWDFRWYLSVPKSFEDFAYSSGPFEEITAVSFSEYVLT 205
Db 170 VHHLDHLQESPEFSDAGAREAMSDSTLIVPREVIERIAASQDWVEIIVAGTLVFEPLVG 229
QY 206 NLLFVP-FMSCAANGMDSTVTFGSAQSDSRHMTIGIEICKLLEQDP-----DNVPIV 260
Db 230 HLAKEAFSRAPMGPDGDTTPAVLASALLDSGRHLE-SVQALVRLVCQDPVHGQDQNAV 288
QY 261 QRWIDKW 267
Db 289 RWLEEW 295

RESULT 6
ID R81469 standard; Protein; 343 AA.

```

AC R81469;
DT 07-AUG-1996 (first entry)
DE Nocardia corallina alkene mono-oxygenase gene product, amcA.
KW Alkene mono-oxygenase; indole; indigo production; biosynthesis;
KW microbial oxidation; dye.
OS Nocardia corallina B-276.
PN J08023988-A.
PD 30-JAN-1996.
PF 08-JUL-1994; 179688.
PR 08-JUL-1994; JP-179688.
PA (NIIH) JAPAN ENERGY CORP.
DR WPI: 96-133426/14.
DR N-PSDB; T17418.
PT Prep. of indigo by a microbiological method - by culturing a
PT microbe having alkene monoxygenase activity to oxidise indole to
PT indigo.
PS Claim 5: Page 6-7: lipp; Japanese.
CC R81469-R81472 are protein products of the Nocardia corallina strain
CC B-276 alkene mono-oxygenase gene. The gene encodes 4 protein products
CC amcA, amcB, amcC and amcD derived from the 3 different reading frames
CC of the operon. The gene is useful for the production of indigo via
CC oxidation of indole. Nocardia corallina can be cultured in a medium
CC contg. indole and will readily oxidise the indole yielding indigo into
CC the culture medium. E. coli may also be transformed with the alkene
CC mono-oxygenase gene and used as above to efficiently produce indigo
CC by microbial oxidation.
SQ Sequence 343 AA;

Query Match 3.8%; Score 109; DB 1; Length 343;
Best Local Similarity 19.4%; Pred. No. 0.0047;
Matches 48; Conservative 39; Mismatches 118; Indels 42; Gaps 6;

QY 44 IKIHWDKWDPPRLTMDAYWKYQGEKKLYAYIDAFTONNAFLGVSDDARYINALKLFL 103
DB 69 VRTSESNAYRPHQLWQRPVYSTCNCQQAALRPVLTMGSA----- 111
QY 104 QGVVPL-----EYLHGFHAGHGHFTGEGARTACQMSIDELRHQYQTE 147
DB 112 -AIIPIWSQILARSYAAPVEYGLFSLAYAYRQAMSDTVQFVQAVDRMRLLODI 170
QY 148 THAMSTYNKF--FNGFHHSNWFDRVWYLSVPKFFEDAYSGPPEFTAVSFSEYVIT 205
DB 171 VHLDDHGESPEFSDAGAREAWMSDSTLVPINIEVIERAASQDWVEILLVAGTLVEPLVG 230
QY 206 NLLFVP-FWGAAYNGDMSTVTFGFSQAQSDSRHMTLGEICIKFLLEQDP-----DNVPIV 260
DB 231 HLAKELEFSRRAPMGDGTTPAVLASALLDSGRHLE-SVQALVRLVCDPVGHDQNAIV 289
QY 261 QRWDKW 267
DB 290 RWIEEW 296

RESULT 7
W12955
ID W12955 standard; protein; 514 AA.
AC W12955;
DT 07-APR-1997 (first entry)
DE Alpha-amylase.
KW Alpha-amylase; detergent; thermal stability; oxidation stability; enzyme;
KW calcium ion dependency; alpha-amylolytic activity; washing composition;
KW textile desizing; papermaking; beer-making; ethanol production;
KW sweetener.
OS Bacillus stearothermophilus.
PN W09623873-A1.
PD 08-AUG-1996.
PF 05-FEB-1996; PF DK00056.
PR 03-FEB-1995; DK-000126.
PR 29-MAR-1995; DK-000336.
PR 29-SEP-1995; DK-001097.
PR 06-OCT-1995; DK-001121.
PA (NOVO) NOVO-NORDISK AS.

PI Bisgard-frantzen H, Borchert T, Svendsen A;
DR WPI: 96-371423/37.
PT Alpha-amylase variants - with improved thermal and oxidation
PT stability and reduced calcium ion dependency
PS disclosure; Page 86-88; l1pp; English.
CC W12955, W12956, W18135 and R81836 represent the parent alpha-amylases
CC used to create the variants of the invention (such as W12098-W12131).
CC This sequence represents the alpha-amylase from Bacillus
CC stearothermophilus. The variants of the invention were created using
CC site directed, or random, mutagenesis of the DNA sequences encoding these
CC parent alpha-amylases. The variants of the invention can have improved
CC thermal stability (such as at temperatures in the range of 40-70 degrees
CC Celsius), and/or oxidation stability, and/or reduced calcium ion
CC dependency. The variants can also have increased alpha-amylolytic
CC activity (especially at pH values in the range of 8.5-10.5), and
CC improved binding of a particular substrate. The variant alpha-amylases
CC also possess improved specificity to a particular substrate, and/or
CC improved specificity with respect to cleavage of substrate, and for textile
CC can be used in detergent and washing compositions, and for papermaking and
CC desizing. The alpha-amylase variants can also be used in beer-making and
CC the production of sweeteners. The variants of the invention can also be used in
CC the production of sweeteners and ethanol from starch.
SQ Sequence 514 AA;

Query Match 3.6%; Score 104; DB 1; Length 514;
Best Local Similarity 18.3%; Pred. No. 0.029;
Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVPFYDYEGI-----KIHWDKW-----VDPFLRTMDAY 63
DB 179 RGIGKANDWEVDTEGNGYDILYADLMDHDEVTLSKWKVYNTNIDGRLDAVKH 238
QY 64 WKYQGEKELKLYAYIDAFTONNAFL--GVSDARYINALKFLQGVTPLEYLAHGFHAGV 121
DB 239 IKF-----SFPDWLSDVRSQTGKPLFTVG-----EYWSY----- 268
QY 122 RHFTGEGARIACOMSIDELRHQYOTETHAMSTYNKFFNGFHHSNQWFDEWVLSVPKSF 181
DB 269 -----DINKLHNYIMKTN--GTMSLFDAPLHN-----KPY 296
QY 182 EDAYSGPPEFTAVSFSEYVITNLLFPVFMGAAYNGDMSTVTFGFSQAQSDSRHMTL 241
DB 297 TASKSGGTFDMKT-----LMTNLM-----KQDFL----- 322
QY 242 GIECIKFLLEQDPDNVPIVQRWDKW-----WRGYRL----- 275
DB 323 ----AVTFVNDHTEPGQALQSWDPWFKPLAYAFILTRQGYPCVGYGYGIPQYNIPS 379
QY 276 -----TLVA-----MMMDYMQPKRVMSW-RE-----SWEM 299
DB 380 LKSKIDPILLIARRDYAYGTQHDYLDHSDITIGWTREGVTEKPGSGLAALITDGGGSKWY 439
QY 300 YAEQNGGALFKDLARYGIREPKGWQDACEGKHISHQAMSTFYGFNAASAFHFWP---T 356
DB 440 VGRQHAGKYDYDLT--GNR-----SDVT-----INSDGWGEF-KVNGGSV-SWVPRKIT 486
QY 357 EDEMGMWLSAKYP--DSFDRYRPRFDHW 382
DB 487 VSTIANSITRPTDEVRWTEPLVAV 514

RESULT 8
W31502
ID W31502 standard; protein; 515 AA.
AC W31502;
DT 08-APR-1998 (first entry)
DE Bacillus sp. alpha amylase.
KW Alpha amylase; hard surface cleaning; dishwashing; laundry.
OS Bacillus sp.
PN W09732961-A2.
PD 12-SEP-1997.
PF 04-MAR-1997; U03635.

PR 07-MAR-1996; WO-U03276.
 PA (PROC) PROCTER & GAMBLE CO.
 PI Baeck AC, Jones LA, Ohtani R, Pramod K, Rai S,
 PI Shewell MS, Ward G;
 DR WPI; 97-457524/42.
 PT Detergent compositions for hard surface cleaning and laundry use -
 PT contains Bacillus derived alpha amylase with improved
 PT thermostability, reduced calcium ion dependency etc.
 PS Claim 1; Page 91; 97pp; English.
 CC The present sequence is a Bacillus sp. alpha amylase with a
 CC specific activity at least 25% higher than that of Termamyl (RTM)
 CC at 25 to 55 degrees C at pH 8 to 10, measured by the Phadebas (RTM)
 CC activity test. It is of use in hard surface cleaning, hand or
 CC machine dishwashing and laundry at a temperature of 10 to 25
 CC degrees C. Improved cleaning, stain removal and fabric care are
 CC obtained by using it at a concentration of 0.00018 to 0.06%
 SQ Sequence 515 AA;

Query Match 3.6%; Score 104; DB 1; Length 515;
 Best Local Similarity 18.3%; Pred. No. 0.029;
 Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVPFYDRYEGI-----KIHMDKW-----VDFRLTMDAY 63
 DB 179 RGIGKAWDEVTENGNDYLMADLMDHPEVVTLSKSGKWVNTNIDGFRDVAKH 238
 QY 64 WKYQGEKKEKLYAVIDAFTQNNAPL--GVSDARYINALKLFLOGVTPLEYLAHGFHVG 121
 DB 239 IKF-----SFFPDWLSVRSQTKPLFTVG-----EYWSY-----268
 QY 122 RHTGEGARIACOMQSIDELRHVQIETHAMSTYKFFNGFHHSNQWFDVWYLSVPKSPF 181
 DB 269 -----DINKLHNYIMKTN--GTMSLFDAPLHN-----KFY 296
 QY 182 EDAYSSGPFELTAVSFSEFVYVLTNLLFVPFMSGAYNGDMSTVTFGSAQSDSRHMTL 241
 DB 297 TASKSGGTDMRT-----LMTNTLM-----KDQPTL-----322
 QY 242 GIECIKFLLEQDPDNPVIVORWIDKWF-----WRGYRL-----275
 DB 323 ---AVTEVDNHDTEPQALQSVDPWFKPLAYAFILTROEGYPCVFGYGIPOYNIPS 379
 QY 276 -----TLVA-----MMMDYMPKRVMSW--RE-----SWEM 299
 DB 380 LKSKIDPLLIARRDAYCTQHDYLDHSDIIGWTRGVTGKPGSGLAALITDGGGSKWY 439
 QY 300 YAEQNGGALFKDLARYGIREPKGWODACGKHISHQAWSTFYGFNAASFHTWVP---T 356
 DB 440 VGKHAGKVFYDLT--GNNR-----SDTVT-----INSDGWGEF--KVNGGSV-SVWVPRKIT 486
 QY 357 EDEMGWLSAKYP--DSFDYRYRPRDHW 382
 DB 487 VSTIAWSITTRPWTDFEVRWTEPRIVAW 514
 RESULT 9
 W48263
 ID W48263 standard; protein; 515 AA.
 AC W48263;
 DT 02-JUL-1998 (first entry)
 DE Bacillus sp. alpha amylase protein #3.
 KW Alpha amylase; stain digestion; detergent; fabric laundry performance.
 OS Bacillus sp.
 PN WO9805748-A1.
 PD 12-FEB-1998.
 PF 01-AUG-1996; U12612.
 PR 01-AUG-1996; WO-U12612.
 PA (PROC) PROCTER & GAMBLE CO.
 PI Baeck AC, Jones LA, Ohtani R, Pramod K, Rai S,
 PI Shewell MS;
 DR WPI; 98-159168/14.
 PT Use of specific alpha-amylase enzymes - in laundry detergent

PT compositions to provide effective cleaning and whitening of dingy
 PT fabrics
 PS Claim 1; Page 74; 82pp; English.
 CC This sequence represents an alpha amylase from Bacillus sp. which is
 CC used in a laundry detergent. The detergent compositions can be used
 CC for boosting fabric laundry performance or for dingy
 CC fabric cleanup.
 SQ Sequence 515 AA;

Query Match 3.6%; Score 104; DB 1; Length 515;
 Best Local Similarity 18.3%; Pred. No. 0.029;
 Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVPFYDRYEGI-----KIHMDKW-----VDFRLTMDAY 63
 DB 179 RGIGKAWDEVTENGNDYLMADLMDHPEVVTLSKSGKWVNTNIDGFRDVAKH 238
 QY 64 WKYQGEKKEKLYAVIDAFTQNNAPL--GVSDARYINALKLFLOGVTPLEYLAHGFHVG 121
 DB 239 IKF-----SFFPDWLSVRSQTKPLFTVG-----EYWSY-----268
 QY 122 RHTGEGARIACOMQSIDELRHVQIETHAMSTYKFFNGFHHSNQWFDVWYLSVPKSPF 181
 DB 269 -----DINKLHNYIMKTN--GTMSLFDAPLHN-----KFY 296
 QY 182 EDAYSSGPFELTAVSFSEFVYVLTNLLFVPFMSGAYNGDMSTVTFGSAQSDSRHMTL 241
 DB 297 TASKSGGTDMRT-----LMTNTLM-----KDQPTL-----322
 QY 242 GIECIKFLLEQDPDNPVIVORWIDKWF-----WRGYRL-----275
 DB 323 ---AVTEVDNHDTEPQALQSVDPWFKPLAYAFILTROEGYPCVFGYGIPOYNIPS 379
 QY 276 -----TLVA-----MMMDYMPKRVMSW--RE-----SWEM 299
 DB 380 LKSKIDPLLIARRDAYCTQHDYLDHSDIIGWTRGVTGKPGSGLAALITDGGGSKWY 439
 QY 300 YAEQNGGALFKDLARYGIREPKGWODACGKHISHQAWSTFYGFNAASFHTWVP---T 356
 DB 440 VGKHAGKVFYDLT--GNNR-----SDTVT-----INSDGWGEF--KVNGGSV-SVWVPRKIT 486
 QY 357 EDEMGWLSAKYP--DSFDYRYRPRDHW 382
 DB 487 VSTIAWSITTRPWTDFEVRWTEPRIVAW 514
 RESULT 10
 W59446
 ID W59446 standard; protein; 526 AA.
 AC W59446;
 DT 13-OCT-1998 (first entry)
 DE Arabidopsis thaliana MLO protein homologue.
 KW MLO; mildew; pathogen; resistance.
 OS Arabidopsis thaliana.
 PN WO9804586-A2.
 PD 05-FEB-1998.
 PF 29-JUL-1997; G02046.
 PR 07-MAR-1997; GB-004789.
 PR 29-JUL-1996; GB-015879.
 PR 30-OCT-1996; GB-022626.
 PA (INNE-) INNES CENT INNOVATIONS LTD JOHN.
 PI Bueschges R, Panstruga R, Schulzelefert PMJ;
 DR WPI; 98-159149/14.
 DR N-PSDB; V35031.
 PT New isolated Mlo gene of barley - used to develop products for the
 PT production of transgenic plants which have increased pathogen
 PT resistance
 PS Claim 11; Fig 15; 150pp; English.
 CC The sequence is that encoded by a homologue of the MLO gene, wild-type
 CC Mlo exerts a negative regulatory function on a pathogen defence response,
 CC such that mutants exhibit a defence response in the absence of pathogen.
 CC Down-regulation or out-competition of Mlo function may be used to

CC stimulate a defence response in transgenic plants conferring increased
 CC pathogen resistance, especially resistance to powdery mildew or rust.
 CC The product can also be used for identifying compounds able to
 CC stimulate a defence response in a plant by interaction with
 CC encoded polypeptide.
 SQ Sequence 526 AA;

Query Match 3.6%; Score 103; DB 1; Length 526;
 Best Local Similarity 22.0%; Pred. No. 0.038;
 Matches 71; Conservative 43; Mismatches 130; Indels 78; Gaps 17;

QY 45 KIHWDKWDVDFRLTMDAYWKYQGE-KEKKLYAVIDAFTONNAFLVGSVDARYI- 96
 DB 184 RIHOWKWEKST-ADKFPDPTALRKRRTVHVNHAFAKEH-FLGIGKDSVILGWT 237

QY 97 -NALKFLQGVPTPLEYLAHR-GFAHVGRRHFTGEGARIACOMSIDELRHYQTETHAMSTY 154
 DB 238 QSEFLKQFVDSYTKSDYVTLRLGF-IMTHCKG-NPKLNFHKYMMRALEDDEKQVVGISWY 294

QY 155 NKFF-----NGFHHSNOWNFDRWYLSVPKSEFFEDAYSSGPFELTAVSFSEYVILT 205
 DB 295 LWIFWVIFLLNVNGWH-----TYFWIAFT--PFALLAVGKLEHVIA 336

QY 206 NLLFVPMFMSGAAYNGDM-----STVTFGFSQSDSESRHMTLGTECIKFLLEODPDNPVIVQ 261
 DB 337 QLAHEVAEKHVAIGDLVVKFSDEHFWFSKQF-----IVLYLHLFILFQNAFIAFF- 388

QY 262 RWIDKFWRGYRLLTLVAMMDYMOFK-----RVMSWRSEWEMYA--EQNGG-----AL 308
 DB 389 ---FWIWTYGFDSICMGQVRYIYVRLVIGVFIQVLCYSILPLVIAVSMQGSFKKAI 444

QY 309 FKDLARYGIREPKGWODACEGK 330
 DB 445 FEENVQVGL--VGWAQKVQK 463

RESULT 11

ID P70579 standard; Protein; 572 AA.
 AC P70579;
 DT 07-MAY-1991 (first entry)
 DE plasmid pTUB613 heat resistant alpha-amylase product.
 KW Thermostable.
 OS Bacillus stearo-thermophilus.
 PN J62104580-A.
 PD 15-MAY-1987.
 PF 30-OCT-1985; 241302.
 PR 30-OCT-1985; JP-241302.
 PA (HIGE-) HIGETA SHOYU KK.
 DR WPI: 87-173697/25.
 DR N-PSDB; N70920.
 PT DNA expressing highly-thermostable enzyme - formed by adding one
 PT more cysteine codon in structural gene
 PS Disclosure; Fig 6; 12pp; Japanese.
 CC The enzyme is modified from the wild type by the inclusion of a
 CC cysteine residue, increasing the thermostability of the alpha-
 CC amylase.
 SQ Sequence 572 AA;

Query Match 3.6%; Score 102; DB 1; Length 572;
 Best Local Similarity 18.1%; Pred. No. 0.054;
 Matches 81; Conservative 44; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVFYDRIYEGT-----KIHWDKW-----VDPFRLTMDAY 63
 DB 236 RGIGKAWDEVDTEGNYDLYADLMDHPEVVTLEKSGWKYVNTWTTNIDGFRDLDAVHK 295

QY 64 WKYQGEKKLYAVIDAFTONNAFL--GYSDARYINALKIFLGQVTPLEYLAHRGAHV 121
 DB 296 IRF-----SFFPDWLSNVRSGKPLFTVG-----EWSY----- 325

QY 122 RHFTGEGARIACOMSIDELRHYQTETHAMSTYKFFNGFHHSNOWNFDRWYLSVPKSEFF 181
 DB 326 -----DINKLHNYNMTN--GTMSLFDAPLHN-----KFY 353

QY 182 EDAYSSGPFELTAVSFSEYVILTLLFPFMSGAAYNGDMSTVTFGFSQSDSESRHMTL 241
 DB 354 TASKSGGTEDMRT-----LMTNTLM-----KDQPTL----- 379

QY 242 GIECIKFLLEQDPDNPVIVQWIDKWF-----WRGYRLL----- 275
 DB 380 ---AVTFVDNHDTEPGALQSWVDPEKFLAYAFILTRQEGYPCVFYGDYGIPOYNIPS 436

QY 276 -----TLVA-----MMMDYMQPKVMSW-RE-----SWEM 299
 DB 437 LKSKIDPLLIARRDYAYGTQHDYLDHSDILGWTREGVTEKPGSLAALITDGPGRKWMY 496

QY 300 YAEQNGGALFKDLARYGIREPKGWODACEGKDHISHQAWSTYGFNAASAFHTWVP---T 356
 DB 497 VGKQHAGKVFYDLT--GNR-----SDTVT-----INSDQWGEF-KVNGGSV-SVVVPRKTT 543

QY 357 EDEMCLWSAKYP--DSFDYRYRPRFDHW 382
 DB 544 VSTIAWSITRPTWDETFVRWIEPRIVAW 571

RESULT 12

ID P70580 standard; Protein; 559 AA.
 AC P70580;
 DT 07-MAY-1991 (first entry)
 DE plasmid pTUB616 heat resistant alpha-amylase product.
 KW Thermostable.
 OS Bacillus stearo-thermophilus.
 PN J62104580-A.
 PD 15-MAY-1987.
 PF 30-OCT-1985; 241302.
 PR 30-OCT-1985; JP-241302.
 PA (HIGE-) HIGETA SHOYU KK.
 DR WPI: 87-173697/25.
 DR N-PSDB; N70926.
 PT DNA expressing highly-thermostable enzyme - formed by adding one
 PT more cysteine codon in structural gene
 PS Disclosure; Fig 7; 12pp; Japanese.
 CC The enzyme is modified from the wild type by the inclusion of a
 CC cysteine residue, increasing the thermostability of the alpha-
 CC amylase.
 SQ Sequence 559 AA;

Query Match 3.4%; Score 99; DB 1; Length 559;
 Best Local Similarity 18.1%; Pred. No. 0.11;
 Matches 81; Conservative 44; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVFYDRIYEGT-----KIHWDKW-----VDPFRLTMDAY 63
 DB 223 RGIGKAWDEVDTEGNYDLYADLMDHPEVVTLEKSGWKYVNTNIDGFRDLDAVHK 282

QY 64 WKYQGEKKLYAVIDAFTONNAFL--GYSDARYINALKIFLGQVTPLEYLAHRGAHV 121
 DB 283 IRF-----SFFPDWLSNVRSGKPLFTVG-----EWSY----- 312

QY 122 RHFTGEGARIACOMSIDELRHYQTETHAMSTYKFFNGFHHSNOWNFDRWYLSVPKSEFF 181
 DB 313 -----DINKLHNYNMTN--GTMSLFDAPLHN-----KFY 340

QY 182 EDAYSSGPFELTAVSFSEYVILTLLFPFMSGAAYNGDMSTVTFGFSQSDSESRHMTL 241
 DB 341 TASKSGGTEDMRT-----LMTNTLM-----KDQPTL----- 366

QY 242 GIECIKFLLEQDPDNPVIVQWIDKWF-----WRGYRLL----- 275
 DB 367 ---AVTFVDNHDTEPGALQSWVDPEKFLAYAFILTRQEGYPCVFYGDYGIPOYNIPS 423

QY 276 -----TLVA-----MMMDYMPKRVMSW-RE-----SWEM 299
 DB 424 LKSKIDPDLIARRDYAYGTHDYLHSDILIGWTRGVIEKPGSLAALITDGPGRKWMY 483
 QY 300 YAEQNGGALFKDLARYGIREPKGWODACEKGDHISHQAWSTFYGFNAASAFHTWTP--T 356
 DB 484 VKQHAGKVFYDLT--GNR-----SDVT-----INSDGWGEF-KVNGGSV-SVWVPRKTT 530
 QY 357 EDEMGWLSAKYP--DSFDYRYPFRDHW 382
 DB 531 VSTIAWSTTRPWTDEFVRWTEPRIVAW 558

RESULT 13
 W49873
 ID W49873 standard; Protein; 684 AA.
 AC W49873;
 DT 21-DEC-1998 (first entry)
 DE Thermotoga maritima MSB8-6GB4 glycosidase.
 KW Glycosidase; MSB8-6GB4; thermostable enzyme; oligosaccharide;
 KW glucose; sugar; baking; textile; detergent; beta-galactosidase.
 OS Thermotoga maritima strain MSB8-6GB4.
 PN W09824799-A1.
 PD 11-JUN-1998
 PF 08-DEC-1997; U22623.
 PR 10-OCT-1997; US-949026.
 PR 06-DEC-1996; US-056916.
 PA (DIVE-) DIVERSA CORP.
 PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
 DR WPI: 98-362407/31.
 DR N-PSDB: V36922.

Glycosidase enzymes from organisms of the genera Staphylothermus,
 Pyrococcus and Thermococcus - for deriving sugar from
 PT oligosaccharides, useful in the e.g. food processing, textile or
 PT baking industries
 PS Claim 1: Fig 16a-c: 92pp; English.
 CC This is the amino acid sequence of glycosidase MSB6-6GB4, deduced
 CC from a polynucleotide (see V36922) of clone 6GB4 of Thermotoga
 CC maritima MSB8 that grows optimally at 85 degC in high salt medium.
 CC The invention provides 18 polynucleotides (see V36907-24) coding
 CC for thermostable glycosidases (see W49858-75) having glucosidase,
 CC alpha-galactosidase, beta-galactosidase, beta-mannosidase,
 CC beta-mannanase, endoglucanase or pullulanase activity. Vectors and
 CC host cells are also claimed. A method is provided for producing
 CC the enzymes by recombinant techniques. A claimed method for
 CC generating glucose from soluble cell oligosaccharides comprises
 CC contacting a sample (selected from dairy products, fruit juice,
 CC detergent, textile, guar gum, animal feed, plant biomass or waste
 CC product) containing oligosaccharides (selected from maltose,
 CC cellobiose, lactose, sucrose, raffinose, stachyose, verbascose,
 CC cellulose, starch, amylose, glycogen, disaccharides, polysaccharides
 CC and pullulan) with one of the claimed glycosidases such that glucose
 CC is produced.
 SQ Sequence 684 AA;

Query Match 3.4%; Score 98.5; DB 1; Length 684;
 Best Local Similarity 18.3%; Pred. No. 0.17;
 Matches 95; Conservative 67; Mismatches 159; Indels 197; Gaps 29;

QY 54 DPFRLTMDAYWY-----OGEKKLYAVIDAFTQ---NNAFGLGVSDARYI--- 96
 DB 49 DLFKIEDREWIYEREFKEDVKGERVDLVFEGVDTLSVDYLVNGVYLGSTEDMFIEYR 108
 QY 97 -----NALKLFQ-----GV-----TPLEYLAHRGFAHVRHFTG-- 126
 DB 109 FQVNVNLKKNHLKVIKSPIRVKPTLQNGYGVLGSPEDI-----RGYIRKAQTSYGM 163
 QY 127 EGARIACQ-----MOSIDELRHQYQETHAMSTYNNKFFNGFHH----- 163
 DB 164 WGARIVTSIWKPVVLEYVYRARIQDSTAYLLELEKDALVR-VNGFVGEGLNIVEYVN 222
 QY 164 -----SNQWFDRWYLVSPKSPFFEDAYSQGFPEFLTAVSFSEFVLTNLLFV 210

DB 223 GEKIGEPVLEKNGEKLPDGVFHLKDVKLWY-----PNNVGKPYLYDFVFLKDL--- 272
 QY 211 PMSGAAAYNGDMSTVTFGSAQSDSRHMTLGIETKFKLLEQDPDNPV-IVORWIDKWFV 269
 DB 273 -----NGEI-----YREEKK-----IGLRKRVIVOEPEDEGKTFIFEINGEKVFA 312
 QY 270 RGYRLTLVMMMDYMPKRVMSW--RESWEMY-----AEON-----GGALFKDLARYG 316
 DB 313 KG-----ANWIPSENILTWLKEEDYEKLVKMARSAANMLRVWGGGIVYEREIFR 362
 QY 317 IREPKG---WQD---AC-EGKDH-----SHQAWSTFYGFNAASAFHTWVPTEDMGW 362
 DB 363 LCDELGIMVWQDFMVALEYDPDLPWFRKLANEERARKIVRKLRYHPSIVLWC-GNNENNW 421
 QY 363 LSAKYPDSFDYRYPFRDHWGEQAR-----AGNREYMKTLPLMC--QVCQIPMLFTEP- 413
 DB 422 -----GFDEWGNMARKVDGINLGNRLYLFDFPEICAEDPSTPYWSPSY 466
 QY 414 ----GNPTKIGARE-----SNYLGKHFHFCSD 436
 DB 467 GGEKANSEKGRHVVYVWSGWMNVENYEKNTGRFISE 504

RESULT 14
 R04111
 ID R04111 standard; peptide; 1306 AA.
 AC R04111;
 DT 07-SEP-1990 (first entry)
 DE Human angiotensin converting enzyme (ACE)
 KW human angiotensin converting enzyme; hypertension; bradykinin.
 OS synthetic.
 FH Key
 PT protein
 FT Location/Qualifiers
 FT 30..1277
 FT /label-mature ACE
 FT /note-derived from pre-ACE by removal of signal peptide"
 FT modified_site 38..38
 FT /label-putative N-glycosylation site
 FT modified_site 54..56
 FT /label-putative N-glycosylation site
 FT modified_site 74..76
 FT /label-putative N-glycosylation site
 FT modified_site 111..113
 FT /label-putative N-glycosylation site
 FT modified_site 146..148
 FT /label-putative N-glycosylation site
 FT modified_site 160..162
 FT /label-putative N-glycosylation site
 FT modified_site 318..320
 FT /label-putative N-glycosylation site
 FT modified_site 445..447
 FT /label-putative N-glycosylation site
 FT modified_site 509..511
 FT /label-putative N-glycosylation site
 FT modified_site 523..525
 FT /label-putative N-glycosylation site
 FT modified_site 677..679
 FT /label-putative N-glycosylation site
 FT modified_site 713..715
 FT /label-putative N-glycosylation site
 FT modified_site 760..762
 FT /label-putative N-glycosylation site
 FT modified_site 942..944
 FT /label-putative N-glycosylation site
 FT modified_site 1191..1193
 FT /label-putative N-glycosylation site
 FT modified_site 1225..1227
 FT /label-putative N-glycosylation site
 FT W09003435-A.
 PN 05-APR-1990.
 PD 27-SEP-1989; F00496.
 PR 27-SEP-1988; FR-012620.
 PA (INRM) Institut National de la Sante et de la Recherche Medicale.

PI Soubrier F, Albenc-Gelas F, Hubert C, Corvol P;
DR WPI: 90-132272/17.
DR N-PSDB; Q04027.
PT New DNA encoding human angiotensin converting enzyme used eg in
PT diagnosis of hypertension, evaluation of enzyme inhibitors
PS Disclosure; p; French.
CC Human angiotensin converting enzyme hydrolyses angiotensin I and kinins.
CC Either intact enzyme or fragments thereof can be used to generate
CC antibodies for diagnostic use. Oligonucleotide probes can also be made
CC which are complementary to the sequence encoding the enzyme.
SQ Sequence 1306 AA;

Query Match 3.3%; Score 94.5; DB 1; Length 1306;
Best Local Similarity 18.8%; Pred. No. 1.2;
Matches 142; Conservative 80; Mismatches 214; Indels 321; Gaps 46;

QY 1 MDPTLKKKLGKLDYAAAMTGLGWETTY---QPMKDV--FPDYREGIKIHWDKW--- 52
Db 428 VSTPEHLHKIGLLDR---VINDTESDINYLKMALEKIAFLPF---GYLVDQW-RWGVE 479

QY 53 ---VDPRLTWDAW---KYQEKERKLYAIDAFTONNAFLGVSDARYINALKLFQ 105
Db 480 SGRTPPSRYNFD-WWYLRTKYQ-----ICPPVTRN-----ETHFDAGAKFHPVN 523

QY 106 VTP-----LEYLAHRCFAHVGRH-----FTGEGARIACOMO----- 136
Db 524 VTYIRYVSFVLQFQHEALCKEAGYEGPLHQCDIYRSYKAGAKLKKVLQAGSSRPQWE 583

QY 137 -----SIDE---LRHYOTETHAMSYNKFNGHHNQNFDRVWYLSVPKSFEDA 184
Db 584 VLKDMVGLDALDAQPLLKYOPVQWLQEQNO-QNG-EVLGWPEYQWHPPLP-----DN 635

QY 185 YSSPPFEFLT---AVSFSEYVLTNLLFVFFMSG-AYNGDMSTVT----- 226
Db 636 YPEG-IDLVTDDEASKFEVEYDRTSOVWVNEAYEAANNYNTITETSKILLQKNMQIA 694

QY 227 ---FGFSAQS-----DESRHMTLIGIE----- 244
Db 695 NHTLKYGTQARKEDVNLQNTIKRIIKYQDLERAAALPAQELLEYKNILLDMETYSVA 754

QY 245 -----CIKFLQPDNDVPIVORWIDK--WFWRGYRLTLIVAMMDYMPKRV--- 291
Db 755 TVCHPNQSCILQ--LEPDLTNVMTATSRKYEDLLWAEQWGRDKAGRAILQFY--PKYVELLN 810

QY 292 -----SWRESWENYA-EQNGGALFKDL-----ARYGIREPKGWQDAC 327
Db 811 QAARLNGYVDAGDSWRSMTYETPSLEQDLERLFOELQPLYLNLHAYVRRALHRYGAO--- 867

QY 328 EGKDHIS-----HQAWSFYG-----FNAASAFHTWVPTED--EMGWSAKY 367
Db 868 ---HINLEGPAPALLGNWQAOTWSNIYDLVVFPPSAPSDT---TEAMLKQGTWPRM 920

QY 368 PDSFDRIYR-----PRDHWG-----EQARAGNR-----FY----- 393
Db 921 FKEADDFTSIGLLFVPPPEF--WNKSMLEKPTDGREVVCHASAWDFYNGDKFRKQCTTV 978

QY 394 -----MKTLPMICOTCOLPMLFTPEGNETKIGARES----- 424
Db 979 NLEDLVVAHEMGHIQFMQKDLPLVALRGANP---GFHEAGDVLALSYSTPKHLHSL 1035

QY 425 NYLGNN-----PHFCSHDCK-DIFDHE--PORYVQAWLPVHQ 458
Db 1036 NLLSSEGSDHEDINFLMKMALDKIAIFPSYLVQDWRWRVFDGSIITKENYQEWWSLRL 1095

QY 459 IHQNCFFP-----DADPGAGEDPLAANLDYAVT 489
Db 1096 KYQGLC-PPVPTQGDGDPGAKFIPSSVPIRYFVS 1131

RESULT 15
W69165
ID W69165 standard; Protein; 1965 AA.

AC W69165;
DT 07-DEC-1998 (first entry)
DE Streptococcus pneumoniae IgA protease.
KW IgA protease; meningitis; pneumonia; antibiotic; vaccine.
OS Streptococcus pneumoniae strain P110.
FH Location/Qualifiers
FT Key
FT Region 96..132
FT /note= "region resembles C-terminal anchor for
FT /note= other Gram-positive surface proteins"
FT Region 96..101
FT /note= "LPNTGS motif"
FT Region 102..125
FT /note= "hydrophobic region"
FT Region 126..132
FT /note= "lysine-rich charged region"
FT Region 187..194
FT /note= "ATP/GTP binding consensus"
FT Region 396..628
FT /note= "hydrophilic region comprising multiple
FT /note= tandem repeats"
FT Peptide 396..415
FT /note= "repeat 1"
FT Peptide 416..436
FT /note= "repeat 2"
FT Peptide 437..453
FT /note= "repeat 3"
FT Peptide 454..470
FT /note= "repeat 4"
FT Peptide 472..483
FT /note= "repeat 5"
FT Peptide 485..501
FT /note= "repeat 6"
FT Peptide 507..523
FT /note= "repeat 7"
FT Peptide 534..546
FT /note= "repeat 8"
FT Peptide 547..560
FT /note= "repeat 9"
FT Peptide 595..607
FT /note= "repeat 10"
FT Peptide 614..628
FT /note= "repeat 11"
FT Binding_site 1605..1609
FT /note= "zinc binding domain (with residue 1629)"
FT Cleavage_site 1629
FT /note= "zinc binding domain (with residues
FT CA2196502-A.
FT 24-MAR-1998.
PD 31-JAN-1997; 196502.
PR 23-SEP-1996; US-026939.
PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.
PA (NEW-) NEW ENGLAND MEDICAL CENT HOSPITALS INC.
PI Gilbert-Rothstein JV, Plaut AG, Weiser JN;
DR WPI: 98-399668/35.
DR N-PSDB; V44592.
PT New DNA encoding Streptococcus pneumoniae IgA protease - and
PT antibiotic screening assay and subunit vaccine etc
PS Claim 11; Fig 4; 66pp; English.
CC This is the amino acid sequence of Streptococcus pneumoniae IgA
CC protease, deduced from an isolated iga gene (see V69165). The
CC invention provides a method of identifying a candidate antibiotic
CC compound that involves incubating the IgA protease with an IgA
CC protein, and comparing IgA protease activity in the presence and
CC absence of a test compound. S. pneumoniae cells or isolated IgA
CC protease enzyme can be used in this method. The antibiotic is
CC used to treat S. pneumoniae or other bacterial infections. The
CC invention also relates to a subunit vaccine comprising an isolated
CC IgA protease polypeptide. This can be used to prevent or reduce
CC the severity of (S. pneumoniae) infections.
SQ Sequence 1965 AA;

Query Match 3.3%; Score 94; DB 1; Length 1965;
Best Local Similarity 21.0%; Pred. No. 2.4;
Matches 97; Conservative 56; Mismatches 193; Indels 116; Gaps 23;

QY 36 FPYDRYEGIK-----IHDWQKWDPERLTM-----DAYWKYQ-----GEKEKK 73
Db 1222 FLYDRDSIVKEVLPDLQKLDYQSDAIRKLTIGISPEVKLTLEYLQCFQSKTKQNLGDSLKK 1281
QY 74 LYAVIDA--FTQNNAFLGVSADARYINALKLPLOGVTPLE-----YLAHRG 116
Db 1282 LLSA-DAGLASHNSATRCYLVDKIKNNKEALLGLTYLERWYNENYGOVNVKDLVMVHPD 1340
QY 117 FAHVGR-----HFTGEGARIAQMSID-----ELRHVQTEHAMST---YNKFFNGF 161
Db 1341 FFRKGNTPDLTLIELGKSGFNLLAKNNVDYGISLASQHGATDLFSTLEHYRKVFLPN 1400
QY 162 HHSNQWF--DRVWVLSVPKSPFEDA-----YSSGPEFLTAVSFSEYVLTNLLF 209
Db 1401 TSNDWFKSETKAYIVFEFKSTIEEVKTKQGLAGTKYSIGVYDRITSATWKYRNVMVPLLT 1460
QY 210 VPMSGRAYNGDMSTVTGFSQAQSDSRHMTLGIECTKFLLEQDPDNNVPIVQRWIDKWF 269
Db 1461 LPERSVFEVIS--TMSLGFAGYDRYTSDHKA--GKALNDFVEENARETAKRQRDHYDW-- 1516
QY 270 RGYRLTLVAMMDYMQPKRYMSRWESWEMAEQNGGALFKDLARYGIREPKGWODACEG 329
Db 1517 --YRILV-----NSQPRKTLFVR--FSLY-----DAYKFGDDTTSG--KATAEA 1554
QY 330 KDHISHQAWSTIFYGFENASAFHTW---VPTEDMGWLSAKYPDSFDRY-YRPRFDH----- 381
Db 1555 KFDSSNFAMKNFFGVPVGNKVYVHGHGAYATGDBGVYVMSYRMLDKHGAINYTHEMTHDSQ 1614
QY 382 -----WGEOARAGNRFRYMKTLPLMLCQTCQIPMLFTEPGNPT 417
Db 1615 DIYLGYYGRRNGLGPEFFAKGL-----LQAP---DQPSDPT 1647

Search completed: September 26, 2000, 20:11:49
Job time: 10373 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:22:40 ; Search time 60.65 seconds
(without alignments)
130,418 Million cell updates/sec

Title: US-09-430-029-5
Perfect score: 2873
Sequence: 1 MDPTTLKKLGLKDRYAAMT.....FDGSEDKNFAWRGQATRN 516

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2409	83.8	519	1	US-08-319-387-4
2	374.5	13.0	500	5	5171684-2
3	370	12.9	501	1	US-08-499-215-4
4	255	8.9	52	3	US-08-917-299-25
5	192	6.7	52	3	US-08-917-299-10
6	152	5.3	327	5	5171684-6
7	109	3.8	342	1	US-08-499-215-2
8	91.5	3.2	548	1	US-08-468-700-37
9	91.5	3.2	548	1	US-08-645-971-5
10	91.5	3.2	548	2	US-08-468-698-35
11	91.5	3.2	548	2	US-08-468-220-35
12	91.5	3.2	548	2	US-08-704-706A-37
13	91.5	3.2	548	3	US-08-890-383-6
14	91.5	3.2	548	3	US-08-914-679A-6
15	91.5	3.2	548	4	PCT-US94-01553A-35
16	91.5	3.2	548	4	PCT-US95-10426-35
17	90.5	3.2	336	3	US-09-095-163-2
18	90	3.1	217	2	US-08-176-414B-3
19	90	3.1	3200	2	US-08-477-451-8
20	89	3.1	868	2	US-08-162-081B-33
21	89	3.1	868	2	US-08-780-872-33
22	88	3.1	549	1	US-08-720-899-6
23	88	3.1	549	1	US-08-459-610-6
24	88	3.1	549	2	US-08-343-804-6
25	88	3.1	549	2	US-08-687-399-6
26	88	3.1	549	2	US-08-600-908A-6
27	88	3.1	549	3	US-08-683-838A-6
28	88	3.1	583	5	5256558-4

29	88	3.1	713	2	US-08-849-212-4	Sequence 4, Appli
30	87	3.0	876	1	US-08-785-071A-2	Sequence 2, Appli
31	87	3.0	876	3	US-09-012-872-2	Sequence 2, Appli
32	86.5	3.0	972	3	US-08-335-844A-24	Sequence 24, Appl
33	85.5	3.0	475	2	US-08-272-255-13	Sequence 13, Appl
34	85.5	3.0	475	4	PCT-US95-08565-13	Sequence 13, Appl
35	84.5	2.9	604	4	PCT-US96-12860-4	Sequence 4, Appli
36	80.5	2.8	365	2	US-08-833-610-7	Sequence 7, Appli
37	80.5	2.8	365	3	US-08-834-033A-17	Sequence 17, Appl
38	80	2.8	52	3	US-08-917-299-26	Sequence 26, Appl
39	80	2.8	976	3	US-08-750-141A-1	Sequence 1, Appli
40	79.5	2.8	529	5	5217865-2	Patent No. 5217865
41	79.5	2.8	604	2	US-08-511-485-6	Sequence 6, Appli
42	79	2.7	866	1	US-08-386-727-8	Sequence 8, Appli
43	79	2.7	866	2	US-08-600-452A-8	Sequence 2, Appli
44	78.5	2.7	438	4	PCT-US95-05922A-2	Sequence 2, Appli
45	78.5	2.7	455	2	US-08-392-625-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-319-387-4
; Sequence 4, Application US/08319387
; Patent No. 5543317
; GENERAL INFORMATION:
; APPLICANT: Shields, Malcolm S.
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,
; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/319,387
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/167,457
; FILING DATE: 15-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,718
; FILING DATE: 02-MAY-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UWF-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 519 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-319-387-4

[illegible]

5 6T01 HFKS CUDWNETS YERRRKKFEETKKYKMPKCKKDRPERVLTIRDYMKMEAEKDRRTHGEL. 64

	Query Match	6.7%;	Score 192;	DB 3;	Length 52;
	Best Local Similarity	70.6%;	Pred.No. 2e-14;		
	Matches 36;	Conservative	7;	Mismatches	8; Indels
QY	97	NALKFLQCVTPELVLAHRGFAVHGRRHFTCEGARACQMOSIDELRHYOTE	147		
			: :	: :	: :
db	1	NALKLELTAVSPLEVCATCGFSEVGHQFGSGARVACMQMIDELRHVQTQ	51		

```
RESULT 6
5171684-6
; PATENT NO. 5171684
; APPLICANT: YEN, KWANG-MU; BLATT, LAWRENCE M.; KARL, MICHAEL R.
; TITLE OF INVENTION: BIOCONVERSIONS CATALYZED BY THE TOULUENE
; MONOOXYGENASE OF PSEUDOMONAS MENDOCINA KR-1
; NUMBER OF SEQUENCES: 41
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/590,374
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 177,631
; FILING DATE: 05-APR-1988
; SEQ ID NO: 6:
; LENGTH: 327
5171684-6

Query Match 5.3%; Score 152; DB 5; Length 327;
Best Local Similarity 24.5%; Pred. No. 1.2e-08;
Matches 64; Conservative 34; Mismatches 131; Indels 32; Gaps 10;

QY 25 WE-TTYQPMKVF-PYDRYEGIKHMDKWDVDFRLTMDAYWKYQGEKEKKLYAVIDAFT 82
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 43 WELSPDPMNLWYKQYRNASPLKHDNDWDAFTDPDOLVRYTNLMQDGESYVQSILFDQFN 102
QY 83 ONNAFLGVSDARYINALKFLQGYTPLEYLAH---RGFAHVGCRHFTGEGARIACOMOSTD 139
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 103 EREHDMQREGWEHTMARC-----SPRLYLFHCLQMSSAYVOOMAPASTISNCCILQATD 158
QY 140 ELR---HYQTETHAMS-TYNKFFNGFHHSNWFDRWYLSVPKSFEDDAYSSGPFELTA 195
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 159 SLRWLTHAYRTHLSLYPDAGLGEHERELWEKEPGQGL-RELME-----KQLTA 209
QY 196 VSFPEYVLNI-----LFVPMGAAYNMGDMSTVTFGSAQSDSRHMTLGIECI 246
: : || : || : || : || : || : || : || : || : || : || : || : ||
DB 210 FDWGEAFVSLNIVVKPMIVESTIFKPLQQQAWENDTLILLIDSQLKDAERHSRWKALV 269
QY 247 KELLEQDPNPVIVQRIWDKW 267
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 270 KHALE-NPDNHAIVIEGWIEKW 289

RESULT 7
US-08-499-215-2
; Sequence 2, Application US/08499215
; Patent No. 5612204
; GENERAL INFORMATION:
; APPLICANT: Saeki, Hisashi
; APPLICANT: Miura, Akira
; TITLE OF INVENTION: BIOLOGICAL DEGRADATIVE
; TITLE OF INVENTION: TREATMENT OF CHLORINE-SUBSTITUTED ETHYLENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue
; CITY: N.W.
; STATE: Washington D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,215
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP Hei-6-179689
; FILING DATE: 08-JUL-1994
```

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-499-215-2

Query Match 3.8%; Score 109; DB 1; Length 342;
Best Local Similarity 19.4%; Pred. No. 0.00084;
Matches 48; Conservative 39; Mismatches 118; Indels 42; Gaps 6;

QY 44 IKIHMDKWDVDFRLTMDAYWKYQGEKEKKLYAVIDAFTONNAFLGVSDARYINALKFL 103
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 68 VRTSESAYRDPHQLWQRPIVSTCMQDQALARLVPLVTMGSA----- 110
QY 104 QGVTEP-----EYLAHGFHVGCRHFTGEGARIACOMQSIDELRHQYTE 147
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 111 -AITPIWSQKILARSYAAMPFVEYGLFLSLAYAVRQAMSDTVQFSVVFQAVDRMLLQDI 169
QY 148 THAMSTYNKF--FNGFHHSNQWFDWYLSVPKSFEDDAYSSGPFELTAVSFSEYVLT 205
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 170 VHLHDHLESPEFSFAGAREAMMSDSTLVPIREVIERTAAASQDWVEILLVAGTLVFEPIVG 229
QY 206 NILFVP-FMSGAAYNMGDMSTVTFGSAQSDSRHMTLGIECIKFLIEQDP----DNVPIV 260
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 230 HLAKELESRRAPMFGDGTTPPAVLASALLDSGRHLE-SVQALVRLVCQDPVHGDQNAV 288
QY 261 QRWIDKW 267
|| : || : || : || : || : || : || : || : || : || : || : || : ||
DB 289 RWIEEW 295

RESULT 8
US-08-468-700-37
; Sequence 37, Application US/08468700
; Patent No. 5736499
; GENERAL INFORMATION:
; APPLICANT: COLIN MITCHINSON
; APPLICANT: CAROL A. REQUADT
; APPLICANT: TRACI H. ROPP
; APPLICANT: LEIF P. SOLHEIM
; TITLE OF INVENTION: MUTANT ALPHA-AMYLASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,700
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 252
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7555
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

;; REFERENCE/DOCKET NUMBER: GC220D1

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7555
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-220-35

Query Match 3.2%; Score 91.5; DB 2; Length 548;
Best Local Similarity 19.1%; Pred. No. 0.16;
Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;

QY 21 RGLG--WETTYQPMKVPFPDYREGI-----KIHWDKW-----VDPFRLTMDAY 63
DB 213 RGIKAWDEVDTEGNGDYLMYADLMDHPVVTELKNGKWYVNTNIDGRL----- 267
QY 64 WKYQGEKELIYAVIDAFQNNAPFLGVSADARYINALKFLQGVTPLEYLAHKGFAHVGRRH 123
DB 268 ---DGLKHIFSEFPD-----WLSYVRSQTKPLFTVG-----EYWSY----- 302
QY 124 FTGEGARIACQMOSIDELRHVOTETHAMSTYKFNFGFHSNQWFDKRWYLSVPKSPFED 183
DB 303 -----DINKLHNYITKTN--GTMSLFADPLHN-----KFTYA 332
QY 184 AYSGPPEFLTAVSFSEFYVLTNLLFVPMGSAAYNGDMST--VTFGSAQSDSRHMTL 241
DB 333 SKSGGAFDMRT-----LMTNTLM-----KDQPTLAVTFVDNHDINPAKRCSH 374
QY 242 GIECIK-----FLEQDPD-----NPIVQWIDKWF--WRGYRLITIVA 279
DB 375 GRPWFKPLAYAFILTRQGYPCVFGYGYIPQYNIPSLKSIDPLLIARRDYAYGT--- 431
QY 280 MMDYMQPKRVMSW-----RESWEMAEQNGGALFKDLARYGI 317
DB 432 -QHDYLDHSDIIGTREGVTEKPGSGLAALITDGAQSKWMYVKGQKAGKVFYDIT--GN 488
QY 318 REPKGWQDACEGKHISHQAWSTFYGFNAASAFHTWVPTEDMGWL-----SAKYPDSFD 372
DB 489 R-----SDTVI-----INSDGWGEF-KVNGGSV-SWVVPKTTVSTIARPIITRPWTGEFV 537
QY 373 RYRPRFDHW 382
DB 538 RWHEPRIVAW 547

RESULT 11
US-08-468-698-35
; Sequence 35, Application US/08468698
; Patent No. 5849549
; GENERAL INFORMATION:
; APPLICANT: Antrim, Richard L.
; APPLICANT: Barnett, Christopher
; APPLICANT: Mitchinson, Colin
; APPLICANT: Power, Scott D.
; APPLICANT: Requaft, Carol
; APPLICANT: Solheim, Leif P.
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,698
; FILING DATE: 06-JUN-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/194,664
; FILING DATE: 10-FEB-94
; APPLICATION NUMBER: 08/016,395
; FILING DATE: 11-FEB-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC220D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7555
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-698-35

Query Match 3.2%; Score 91.5; DB 2; Length 548;
Best Local Similarity 19.1%; Pred. No. 0.16;
Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;

QY 21 RGLG--WETTYQPMKVPFPDYREGI-----KIHWDKW-----VDPFRLTMDAY 63
DB 213 RGIKAWDEVDTEGNGDYLMYADLMDHPVVTELKNGKWYVNTNIDGRL----- 267
QY 64 WKYQGEKELIYAVIDAFQNNAPFLGVSADARYINALKFLQGVTPLEYLAHKGFAHVGRRH 123
DB 268 ---DGLKHIFSEFPD-----WLSYVRSQTKPLFTVG-----EYWSY----- 302
QY 124 FTGEGARIACQMOSIDELRHVOTETHAMSTYKFNFGFHSNQWFDKRWYLSVPKSPFED 183
DB 303 -----DINKLHNYITKTN--GTMSLFADPLHN-----KFTYA 332
QY 184 AYSGPPEFLTAVSFSEFYVLTNLLFVPMGSAAYNGDMST--VTFGSAQSDSRHMTL 241
DB 333 SKSGGAFDMRT-----LMTNTLM-----KDQPTLAVTFVDNHDINPAKRCSH 374
QY 242 GIECIK-----FLEQDPD-----NPIVQWIDKWF--WRGYRLITIVA 279
DB 375 GRPWFKPLAYAFILTRQGYPCVFGYGYIPQYNIPSLKSIDPLLIARRDYAYGT--- 431
QY 280 MMDYMQPKRVMSW-----RESWEMAEQNGGALFKDLARYGI 317
DB 432 -QHDYLDHSDIIGTREGVTEKPGSGLAALITDGAQSKWMYVKGQKAGKVFYDIT--GN 488
QY 318 REPKGWQDACEGKHISHQAWSTFYGFNAASAFHTWVPTEDMGWL-----SAKYPDSFD 372
DB 489 R-----SDTVI-----INSDGWGEF-KVNGGSV-SWVVPKTTVSTIARPIITRPWTGEFV 537
QY 373 RYRPRFDHW 382
DB 538 RWHEPRIVAW 547

RESULT 12
US-08-704-706A-37
; Sequence 37, Application US/08704706A
; Patent No. 5958739
; GENERAL INFORMATION:
; APPLICANT: GENENCOR INTERNATIONAL, INC.
; APPLICANT: COLIN MITCHINSON

QY 373 RYRPRFDHW 382
DB 538 RWHEPRLVAV 547

RESULT 14
US-08-914-679A-6
; Sequence 6, Application US/08914679A
; Patent No. 6080568
; GENERAL INFORMATION:
; APPLICANT: Anthony G. Day
; APPLICANT: Barbara A. Swanson
; TITLE OF INVENTION: MUTANT ALPHA-AMYLASE COMPRISING MODIFICATION
; TITLE OF INVENTION: AT RESIDUES CORRESPONDING TO A210, HA05 AND/OR
; TITLE OF INVENTION: T412 IN BACILLUS LICHENIFORMIS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,679A
; FILING DATE: To Be Assigned
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 846-7555
; TELEFAX: (650) 845-6504
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-914-679A-6

Query Match 3.2%; Score 91.5; DB 3; Length 548;
Best Local Similarity 19.1%; Pred. No. 0.16;
Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;

QY 21 RGLG--WETTYQPMKVPFYDRYEGI-----KIHWDKW-----VDPRLTMDAY 63
DB 213 RGIGKAWDEVDTEGNYDLYMADLMDHDPVTELNKNGKQWYVNTNIDGFR-----267
QY 64 WKYQGEKKLYAVIDAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHGFVGRH 123
DB 268 ---DGLKHIFSPFD-----WLSYVRSQTGKPLFTVG-----EWSY-----302
QY 124 FTGEGARIACQMSIDELRHVQETHTAMSTYKFNFGHNSQWDFRWVILSVKSPFED 183
DB 303 -----DINKLHNYITKTN--GTMSLFDAPLHN-----KEYTA 332
QY 184 AYSGGPFELTAVSFSEYVLTNLLFVPMGAAAYNGDMST--VTFGFSQASDESRLMTL 241
DB 333 SKSGGAFDMRT-----LMTNTLM-----KDQPTLAVTFVDNHDTPAKRCSH 374
QY 242 GIECIK-----FILEQDPD-----NVPVQRWIDKWF--WRGVRLLTVA 279
DB 375 GRPWFKPLAYAFILTRQEGYPCVFGYDYGIPQYNIPSLKSKIDPLLIARDYAYGT---431

QY 280 MMDYMQPKRVMSW-----RESWEMYAEQNGGALFKDLARYGI 317
DB 432 -QHDYLDHSDIIGTREGYTEKPGSLAALLTDAGRSKMWYVKGQKAGKVFYDLT--GN 488
QY 318 REPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTVWVPTDEMGWL-----SAKYPDSD 372
DB 489 R-----SDTWT-----INSDGWGEF-KVNGGSV-SVWVPRKTTVTSTIARPTITRPTWGEFV 537
QY 373 RYRPRFDHW 382
DB 538 RWHEPRLVAV 547

RESULT 15
PCT-US94-01553A-35
; Sequence 35, Application PCT/US9401553A
; GENERAL INFORMATION:
; APPLICANT: GENECOR INTERNATIONAL, INC.
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genecor International, Inc.
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01553A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC220-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SFO ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-01553A-35

Query Match 3.2%; Score 91.5; DB 4; Length 548;
Best Local Similarity 19.1%; Pred. No. 0.16;
Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;

QY 21 RGLG--WETTYQPMKVPFYDRYEGI-----KIHWDKW-----VDPRLTMDAY 63
DB 213 RGIGKAWDEVDTEGNYDLYMADLMDHDPVTELNKNGKQWYVNTNIDGFR-----267
QY 64 WKYQGEKKLYAVIDAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHGFVGRH 123
DB 268 ---DGLKHIFSPFD-----WLSYVRSQTGKPLFTVG-----EWSY-----302
QY 124 FTGEGARIACQMSIDELRHVQETHTAMSTYKFNFGHNSQWDFRWVILSVKSPFED 183
DB 303 -----DINKLHNYITKTN--GTMSLFDAPLHN-----KEYTA 332
QY 184 AYSGGPFELTAVSFSEYVLTNLLFVPMGAAAYNGDMST--VTFGFSQASDESRLMTL 241
DB 333 SKSGGAFDMRT-----LMTNTLM-----KDQPTLAVTFVDNHDTPAKRCSH 374

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:11:49 ; Search time 75.19 Seconds
(without alignments)
37.172 Million cell updates/sec

Title: US-09-430-029-6
Perfect score: 666
Sequence: 1 MAVTALKPYDPVKDAVEKE.....HKDLISPTPELDGLGASG 118

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	370	55.6	93	W06803	Toluene ortho-mono
2	341.5	51.3	119	W98974	Alcaligenes sp. pr
3	78.5	11.8	495	R41362	TSS1. Trehalose sy
4	78.5	11.8	495	W37441	Trehalose-6-phosph
5	78.5	11.8	495	W37427	Yeast trehalose-6-
6	74.5	11.2	666	W64590	Human SRCR protein
7	74.5	11.2	1785	W64591	Human SRCR protein
8	72.5	10.9	495	R44923	GSL1 glucose sensi
9	72.5	10.9	495	W49026	Saccharomyces cere
10	71.5	10.7	273	W23277	Bordetella pertuss
11	66.5	10.0	345	R14408	Nuclear factor C/E
12	65	9.8	261	R63805	Polyhydroxyalkanoa
13	65	9.8	422	R99090	Human interleukin-
14	65	9.8	423	R92814	Human interleukin-
15	65	9.8	491	W52826	Human chorionecto
16	65	9.8	597	W52821	Human chorionecto
17	65	9.8	834	W52820	Human chorionecto
18	64	9.6	1274	W82953	Human PRCC-TFE3 co
19	63.5	9.5	524	W32096	Human ALP. New nuc
20	63	9.5	105	W74887	Miniature swine re
21	62.5	9.4	876	Y04883	Human secreted pro
22	62.5	9.4	886	Y04884	Mycobacterium spec
23	62.5	9.4	1450	W30751	Rat phospholipase-
24	62	9.3	622	R92702	Rice HY4 gene prod
25	61.5	9.2	596	W48785	Thyroid peroxidase
26	61.5	9.2	917	R64974	Pol region gene pr
27	60.5	9.1	58	Y12746	Human 5' EST seque
28	60.5	9.1	376	W78490	T. thermophilus ga
29	60.5	9.1	376	W78490	Thermus thermophil
30	60.5	9.1	392	W73765	M. tuberculosis an
31	60.5	9.1	392	W73655	M. tuberculosis an
32	60.5	9.1	552	W71636	Omega-cyclohexane
33	60.5	9.1	810	R04574	Derived amino acid

34	60	9.0	123	1	W78326	Fragment of human
35	60	9.0	467	1	W76426	Glucocorticoid ind
36	60	9.0	467	1	Y05284	EGF-like homologue
37	60	9.0	467	1	Y13346	Amino acid sequenc
38	59.5	8.9	304	1	W13405	Eupenicillium bref
39	59.5	8.9	524	1	W32091	Porcine retrovirus
40	59.5	8.9	681	1	R92701	Arabidopsis blue 1
41	59.5	8.9	810	1	R11255	Murine IL-4 recept
42	59.5	8.9	810	1	W13498	Murine interleukin
43	59.5	8.9	810	1	W48313	Mouse interleukin-
44	59.5	8.9	810	1	W60669	Mouse interleukin-
45	59.5	8.9	810	1	W73471	Mouse interleukin-

ALIGNMENTS

RESULT 1

W06803

ID W06803 standard; Protein; 93 AA.

AC W06803;

DE 29-JAN-1997 (first entry)

DE Toluene ortho-monoxygenase subunit tomA4.

KW tom; PTOM; self-transmissible; constitutive; bioreactor; pollutant;

KW breakdown; trichloroethylene; TCE; degradation.

OS Pseudomonas cepacia strain PRL-23.

PN US5543317-A.

PD 06-AUG-1996.

PF 02-MAY-1991; 694718.

PR 02-MAY-1991; US-694718.

PR 15-DEC-1993; US-167457.

PR 06-OCT-1994; US-319387.

PA (FRAN/) FRANCESCONI S C.

PA (SHE/) SHIELDS M S.

PI Francesconi SC, Shields MS;

DI WPI; 96-370640/37.

PT Microorganisms transformed with P. cepacia PRL-23 Tom enzyme gene -
are useful for degradation of chloro:aliphatic cpds. and aromatics
PS Claim 1; Column 33-34; 25pp; English.

CC The present sequence is that of toluene ortho-monoxygenase subunit
tomA4 encoded by T44457, isolated from Pseudomonas cepacia strain
PRL-23. The Tom gene is present on a large self-transmissible plasmid
denoted PTOM. The enzyme is capable of degrading trichloroethylene
(TCE), a hazardous pollutant. The PTOM plasmid is transmissible and
expressable in other bacteria, thus many bacteria can be genetically
altered to constitutively degrade TCE, esp. in bioreactors or
TCE-contaminated environments. P. cepacia PRL-23 contg. PTOM does not
need exogenous chemical inducers and is capable of functioning under a
diverse set of conditions. Also it does not require an inducer that is a
co-substrate for the enzyme required to break down TCE, in effect, it is
not subject to competitive inhibition.
SQ Sequence 93 AA;

Query Match

Best Local Similarity 55.6%; Score 370; DB 1; Length 93;

Matches 63; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

QY 1 MAVTALKPYDPVKDAVEKEFPAPILYVCWENHLMFPAPFCLPLPMDPFAGALAGDVLPPV 60
|:||||| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

Db 1 MSVVALPKYPKPAARENFPAPILFICWEDHLLFAAPVALPLPFDLFGALCTQVLFGT 60
|:||||| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

QY 61 YGYHPDFAKIDWDVWFRSGEPWAPDPAKSLA 93
|:||||| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

Db 61 YGYHPDFSKIDWSQVQWFKSGQPHWPDPAKSLA 93
|:||||| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

RESULT 2

W98974

ID W98974 standard; Protein; 119 AA.

AC W98974;

DT 10-MAY-1999 (first entry)

DE Alcaligenes sp. protein Foxe.

PT by specific ribozyme or stimulator of trehalose-6-phosphate
PT production, particularly for in vitro or in vivo insulin production
PS Claim 13; Page 152-154; 263pp; English.
CC A mammalian cell has been developed comprising an effective amount of a
CC low Km hexokinase-inhibitor selected from: (a) an agent that stimulates
CC production of trehalose-6-phosphate (T6P); or (b) a ribozyme specific
CC for low Km hexokinase. The present sequence represents yeast
CC trehalose-6-phosphate synthase. The cells are particularly used to
CC produce insulin, in response to glucose or other secretagogues, either
CC in vitro or in vivo (for treating diabetes), but may also be used to
CC produce many other therapeutic proteins, e.g. from a gene under control
CC of the insulin promoter which is therefore responsive to glucose.
CC Growth-inhibited cells (having altered hexokinase activity) are used to
CC produce proteins, e.g. insulin or antibodies, in vivo or in vitro.
CC Reduction in low Km hexokinase activity provides cells in which insulin
CC secretion is induced at glucose concentrations closer to the normal
CC range than in the parent cell (nearly homeostatic secretion). Implanted
CC cells of reduced low Km hexokinase activity are expected to survive
CC longer in the host.
SQ Sequence 495 AA;

Query Match 11.8%; Score 78.5; DB 1; Length 495;
Best Local Similarity 24.7%; Pred. No. 0.59; Mismatches 23; Gaps 4;
Matches 23; Conservative 19; Indels 28; Indels 23; Gaps 4;
QY 14 KDAAKFFAPLLYVWENHLMFPAPFCLPLPDPMPFAGLVPPVGYGHPDFAKIDWD 73
DB 80 KDLLEKFNAPVIFLSD-----IADLYNGFSNLIWPLRPHYHP-GEINFD 124
QY 74 RVENF---RSGEWPAPDPAKSLAGNGLGHKDLI 103
DB 125 ENAFPGYNEANQTFNTEAKT-----MHNNDLI 152

RESULT 6
W64590
ID W64590 standard; Protein; 666 AA.
AC W64590;
DE 23-OCT-1998 (first entry)
DE Human SRCR protein fragment.
KW Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;
KW nervous system; medullo-blastoma; glioma; breast; detection;
KW autoantibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 3
FT /label= unknown
FN W09830687-A2.
PD 16-JUL-1998.
PF 09-JAN-1998; D00096.
PR 18-JUL-1997; DE-030997.
PR 09-JAN-1997; DE-000519.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI Mollenhauer J, Poustka A;
DR WPI: 98-399136/34.
DR N-PSDB; V49651.
PT Proteins containing scavenger receptor, cysteine rich domain -
PT useful for diagnosis and treatment of tumours
PS Claim 1; Fig 1; 54pp; German.
CC This sequence represents a fragment of a human protein which contains a
CC SRCR (scavenger receptor, cysteine-rich) domain. The gene and encoded
CC protein can be used to diagnose or treat tumours, particularly of the
CC nervous system (medullo-blastoma or glioma) or breast. The DNA sequence
CC and probes derived from it, are used to identify genes that express
CC SRCR-domain containing proteins, to determine the form in which these
CC proteins exist and to assess the significance of individual forms on
CC cellular properties. The protein can be used to detect the presence of
CC autoantibodies, and Ab which regulate its expression.
SQ Sequence 666 AA;

Query Match 11.2%; Score 74.5; DB 1; Length 666;

Best Local Similarity 38.9%; Pred. No. 2.3;
Matches 21; Conservative 6; Mismatches 20; Indels 7; Gaps 3;

QY 27 VCVENHLMFPAPFCLPLPDPMPF---GAL---AGDVLPVY-GYHPDFAKIDWD 73
DB 405 ICSGNHLSTPAPFLNITRPNTDYSQGFSLQSPGDFSSPFGNYPNNAKCVD 458

RESULT 7
W64591
ID W64591 standard; Protein; 1785 AA.
AC W64591;
DE 23-OCT-1998 (first entry)
DE Human SRCR protein.
KW Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;
KW nervous system; medullo-blastoma; glioma; breast; detection;
KW autoantibody; ss.
OS Homo sapiens.
PN W09830687-A2.
PD 16-JUL-1998.
PF 09-JAN-1998; D00096.
PR 18-JUL-1997; DE-030997.
PR 09-JAN-1997; DE-000519.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI Mollenhauer J, Poustka A;
DR WPI: 98-399136/34.
DR N-PSDB; V49652.
PT Proteins containing scavenger receptor, cysteine rich domain -
PT useful for diagnosis and treatment of tumours
PS Claim 2; Fig 2; 54pp; German.
CC This sequence represents a human protein which contains a SRCR (scavenger
CC receptor, cysteine-rich) domain. The gene and encoded protein can be used
CC to diagnose or treat tumours, particularly of the nervous system
CC (medullo-blastoma or glioma) or breast. The DNA sequence and probes
CC derived from it, are used to identify genes that express SRCR-domain
CC containing proteins, to determine the form in which these proteins exist
CC and to assess the significance of individual forms on cellular
CC properties. The protein can be used to detect the presence of
CC autoantibodies and antibodies which regulate its expression.
SQ Sequence 1785 AA;

Query Match 11.2%; Score 74.5; DB 1; Length 1785;
Best Local Similarity 38.9%; Pred. No. 7.5;
Matches 21; Conservative 6; Mismatches 20; Indels 7; Gaps 3;

QY 27 VCVENHLMFPAPFCLPLPDPMPF---GAL---AGDVLPVY-GYHPDFAKIDWD 73
DB 1356 ICSGNHLSTPAPFLNITRPNTDYSQGFSLQSPGDFSSPFGNYPNNAKCVD 1409

RESULT 8
R44923
ID R44923 standard; Protein; 495 AA.
AC R44923;
DE 05-JUL-1994 (first entry)
DE GGS1 glucose sensing protein.
KW Glucose; sensing; cell signalling; yeast; Saccharomyces cerevisiae;
KW bread; breadmaking; fermentation; alcohol; beverage; brewing;
KW drying; osmotic shock; freezing; sugar metabolism.
OS Saccharomyces cerevisiae.
PN EP-577915-A.
PD 12-JAN-1994.
PF 09-JUL-1992; 870102.
PR 09-JUL-1992; EP-870102.
PA (ALGI-) ALGIST-BRUGGEMAN NV.
PI Holmann S, Thevelein J, Van Dijck P;
DR WPI: 94-009855/02.
DR N-PSDB; Q54385.
PT Transformed yeast with increased stress resistance or
PT fermentation capacity - has modification in general glucose
PT sensor system, partic. for bread-making, but also prodn. of
PT alcohol or foreign proteins

PS Disclosure: Figure 1; 26pp; French.
 CC Yeast strains transformed with an altered GGS1 gene are resistant to
 CC stress and/or have an altered sugar metabolism. The transformed
 CC yeast can be used in industry for the production of bread, alcohol
 CC and beverages and has resistance to at least one of drying, osmotic
 CC shock (especially in sugar containing doughs) and freezing.
 CC Preferably the yeast also has higher trehalose content with delayed
 CC trehalose mobilisation.
 SQ Sequence 495 AA;

Query Match 10.9%; Score 72.5; DB 1; Length 495;
 Best Local Similarity 23.7%; Pred. No. 2.7;
 Matches 22; Conservative 19; Mismatches 29; Indels 23; Gaps 4;
 QY 14 KDAVEKFPAPLLYVCWENHLMFPAPCLPLPDPMPFGALAGDVLPVYGYHDPFAKIDWD 73
 DB 80 KDLLEKFNAPVPLSDE-----IADLHYNGFSNLSILWPLFHYHP--GEINFD 124
 QY 74 RVEWFP---RSGEWPAPDPAKSIAGNGLGHKDLI 103
 DB 125 ENAWLAYNEANQTFNTEIAKT-----MHNNDLI 152

RESULT 9
 W49026
 ID W49026 standard; protein; 495 AA.
 AC W49026;
 DT 07-OCT-1998 (first entry)
 DE Saccharomyces cerevisiae trehalose synthase.
 KW Trehalose synthase; stabilising agent; prokaryotic cell preservation;
 KW prophylaxis; live bacterial vaccine; vaginal infection;
 KW urinary tract infection; live bacterial pharmaceutical.
 OS Saccharomyces cerevisiae.
 PN WO9824882-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997; G03375.
 PR 05-DEC-1996; US-032423.
 FA (QUAD-) QUADRANT HOLDINGS CAMBRIDGE LTD.
 PI Colaco C, Dhaliwal KS, Roser BJ, Tunnaccliffe AG,
 PI Welsh DT;
 DR WPI: 98-333305/29.
 PT Preservation of prokaryotic cells - by increasing intracellular
 PT trehalose concentration, mixing with a stabilising agent, and drying
 PT to produce a glass form of the stabilising agent

PS Disclosure: Pages 39-40; 64pp; English.
 CC The present sequence represents the Saccharomyces cerevisiae trehalose
 CC synthase. The invention provides methods of preserving prokaryotic
 CC cells which involves inducing the intracellular trehalose production
 CC of the prokaryotic cell to an amount which effectively allows increased
 CC storage stability and drying of the cells in the presence of a
 CC stabilising agent. Therefore, prokaryotic cells known to synthesise
 CC trehalose, such as S. cerevisiae, can be used to illustrate the methods
 CC of the invention. Bacteria stabilised by these methods showed less than
 CC 10% loss of viability on storage even after being stored at temperatures
 CC up to at least 37 degrees centigrade for as long as 6 weeks. The
 CC invention claims the methods to be useful for producing dried, stable
 CC prokaryotic cells such as bacteria useful for pharmacological treatment,
 CC prophylaxis, agricultural and industrial applications. The invention
 CC also claims the methods to be useful for producing live bacterial
 CC vaccines in a dry stable form, live bacterial neutraceuticals in a dry
 CC stable form and other live bacterial pharmaceuticals in a dry stable
 CC form, e.g. for treatment of vaginal or urinary tract infections.
 SQ Sequence 495 AA;

Query Match 10.9%; Score 72.5; DB 1; Length 495;
 Best Local Similarity 23.7%; Pred. No. 2.7;
 Matches 22; Conservative 19; Mismatches 29; Indels 23; Gaps 4;
 QY 14 KDAVEKFPAPLLYVCWENHLMFPAPCLPLPDPMPFGALAGDVLPVYGYHDPFAKIDWD 73
 DB 80 KDLLEKFNAPVPLSDE-----IADLHYNGFSNLSILWPLFHYHP--GEINFD 124

QY 74 RVEWFP---RSGEWPAPDPAKSIAGNGLGHKDLI 103
 DB 125 ENAWLAYNEANQTFNTEIAKT-----MHNNDLI 152
 RESULT 10
 W23277
 ID W23277 standard; Protein; 273 AA.
 AC W23277; 1997 (first entry)
 DT 03-NOV-1997 (first entry)
 DE Bordetella pertussis ptSE.
 KW Holotoxin; ptx; pts; gene; extracellular export; ptSE; secretion;
 KW whooping cough; vaccine.
 OS Bordetella pertussis.
 PN US5643747-A.
 PD 01-JUL-1997.
 PF 15-MAR-1993; 031619.
 PR 31-MAR-1994; US-221750.
 PR 15-MAR-1993; US-031619.
 PA (AMCY) AMERICAN CYANAMID CO.
 PI Baker SM, Deich RA;
 DR WPI: 97-350237/32.
 N-PSDB; T84745.
 DR Production of secreted pertussis holotoxin - by culturing
 PT microorganisms transformed with Bordetella pertussis ptx operon and
 PT new pts secretion sequences
 PS Claim 1; Columns 45-48; 3pp; English.
 CC The present sequence is the Bordetella pertussis ptSE, which is
 CC 1 of the proteins encoded by the pts gene that facilitate the
 CC extracellular export of the pertussis holotoxin (ptx), namely
 CC ptSAB, ptSC, ptSE, ptSF, and ptSG. A secreted ptx can be
 CC produced by culturing an E. coli or Bordetella host transformed
 CC with 1 or more vectors, comprising the ptx and pts genes
 CC operatively linked to appropriate expression control sequences. The
 CC secreted ptx can be used in the production of whooping cough
 CC vaccines.
 SQ Sequence 273 AA;

Query Match 10.7%; Score 71.5; DB 1; Length 273;
 Best Local Similarity 21.6%; Pred. No. 1.8;
 Matches 38; Conservative 20; Mismatches 47; Indels 71; Gaps 8;
 QY 3 VIALKPYDPKDAVEKFPAPLLYVCWENHLM-----WNTNLLVKTDRLYDFDLVLASADAATPQALQRS 139
 DB 83 VLIYKAKSFPAQSPQAQPEPGL---WNTNLLVKTDRLYDFDLVLASADAATPQALQRS 139
 QY 35 -----FP-----APFCLPLPDPMPFGAL-----AGDVLPVYGYHDPFAKI 70
 DB 140 RMAYRLQFRYPAPQAASRASPVPAPAPAGALNRRYAMOVGVGSDGIAPTAAY-----192
 QY 71 DMDRVEW---FRSGEPW-----APDPAKSIAGNGLGHKDLISFR--TPGLDGLGAS 117
 DB 193 DDGRHTWLTFRGQPPFAVFAVAPDGTETLVNLHIDNQSLVHRVAPVLMRSGAS 248

RESULT 11
 R14408
 ID R14408 standard; Protein; 345 AA.
 AC R14408;
 DT 11-FEB-1992 (first entry)
 DE Nuclear factor C/EBP2.
 KW Interleukin 6; IL-6; transcription; hematinic.
 OS Homo sapiens.
 PN J03236782-A.
 PD 22-OCT-1991.
 PF 17-OCT-1990.
 PR 25-DEC-1989; JP-336868.
 PR 17-OCT-1990; JP-279650.
 PA (CHUZ-) CHUZO KISH IMOTO.
 DR WPI: 91-35113/48.
 DR N-PSDB; Q15015.

This Page Blank (uspto)